

**ANALYSE ET RECONSTRUCTION DE FAISCEAUX DE  
LA MATIÈRE BLANCHE**

par

François Rheault

Thèse présentée au Département d'informatique  
en vue de l'obtention du grade de philosophiæ doctor (Ph.D.)

FACULTÉ DES SCIENCES  
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, juillet 2020

*Jury*

Professeur Maxime Descoteaux  
Directeur de recherche  
Département d'informatique  
Faculté de sciences Université de Sherbrooke

Professeur Louis Collins  
Co-directeur de recherche  
McConnell Brain Imaging Centre  
Montreal Neurological Institute  
McGill University

Professeur Kevin Whittingstall  
Évaluateur interne  
Département de radiologie diagnostique  
Faculté de médecine et des sciences de la santé  
Université de Sherbrooke

Professeur Bratislav Misic  
Évaluateur externe  
Department of Neurology and Neurosurgery  
Montreal Neurological Institute  
McGill University

Professeur Pierre-Marc Jodoin  
Président rapporteur  
Département d'informatique  
Faculté de sciences  
Université de Sherbrooke

# Sommaire

Les mystères du cerveau sont au centre de l’imaginaire populaire, la complexité de cet organe a toujours été un défi de taille pour son étude. Les avancées récentes dans le domaine de l’imagerie par résonance magnétique de diffusion permettent de mieux modéliser sa connectivité. Les milliards d’axones, qui remplissent la matière blanche et permettent la communication entre les différentes régions du cerveau, sont cruciaux à tous les processus cognitifs.

Cette thèse s’intéresse plus particulièrement à la reconstruction et l’analyse de faisceaux, l’étude de leurs morphologies et à repousser les limites des algorithmes de la tractographie. L’objectif de celle-ci est de procurer des outils permettant d’améliorer la modélisation de la connectivité et de faciliter les applications cliniques qui pourraient en découler dans le futur.

**Mots-clés :** matière blanche, IRM de diffusion, modèle local, tractographie, segmentation de faisceaux, dissection virtuelle, connectivité structurelle, connectome

“I don’t have friends. I got family.”

---

– Dominic Toretto, Furious Seven.



# Remerciements

Mon aventure à l'Université de Sherbrooke aura duré un peu plus de 8 ans. De mon baccalauréat en sciences de l'image et médias numériques, un programme inconnu rempli de programmation qui à l'époque m'effrayait plus que les mathématiques. Jusqu'au doctorat en informatique, programme où les mathématiques m'effraient plus que jamais. Sans le support de mes amis et collègues ainsi que ma famille, cette expérience aurait très bien pu être impossible à achever en restant sain d'esprit.

Ma soif de voyage a été assouvie grâce à Maxime qui n'a pas hésité à me laisser partir à l'étranger dès ma deuxième session de maîtrise et qui m'a encouragé à collaborer (très loin de Sherbrooke) tout au long de mon doctorat. Merci à Flavio, Louis, Laurent, Emmanuel et Maria de m'avoir accueilli dans leur laboratoire malgré la complexité bureaucratique qui en découlait.

Merci à ma famille, mes parents Lily et Laurent et mon frère Benoit, de m'avoir encouragé dans mes études, de ne pas m'avoir demandé trop souvent «*Quand est-ce que tu termines ?*». Un grand merci à mes parents pour m'avoir *autorisé* à voyager à l'époque des Cadets. Sans ces expériences, le cheminement vers les études graduées n'aurait pas été autant attrayant.

Une mention toute spéciale à Jo, Philippe, Jasmeen et Ann-Marie pour les milliers d'heures au Gym ou dans le Mont-Bellevue qui m'ont permis de rester au sommet de ma forme et d'évacuer un peu de stress.

## REMERCIEMENTS

A special word to Jasmeen for the amazing summer we spent in Europe. Together we discovered the joy of surfing in Portugal, wasted too much money in Geneva, drank with Mic in Prague, got stuck in Paris, practiced our kayaking/biking/quadding skills, celebrated Bastille Day on the banks of the Garonne and visited St-Emilion with Mic and Jo. This was truly an amazing summer, one I will remember above all the others.

Sans l'ambiance exceptionnelle du labo créée avant tout par une gang d'amis, la tolérance de tout le monde envers mes goûts musicaux autant excellents qu'exécrables, votre aide pour choisir les meilleures couleurs pour mes figures et m'entendre chialer contre des ignares sur les *interwebs* je n'aurais jamais complété un doctorat. Merci pour la centaine d'heures de *Parks&Recreation*, *Fast&Furious*, *Die Hard* ou *Zombieland*, lasertag, ski/snow au Mont Orford, au 5@8 ou au Siboire qu'on a investies ensemble pour du «*teambuilding*».

Cette thèse n'aurait pas été possible sans un bon standard de programmation, c'est donc crucial de remercier JC pour son support constant, ses réponses toujours utiles et son aide essentielle pour gérer des données dans le mauvais espace, des librairies pas compatibles ou des commandes qui s'interrompent sur Mammouth. Une partie de ma contribution à la recherche va survivre dans *Scilpy* grâce à toi.

Un remerciement final pour Maxime qui a définitivement éveillé ma curiosité pour la recherche scientifique dès le premier cours qu'il nous a enseigné. Mon premier *A* à l'université m'avait à l'époque remonté le moral, mais depuis j'ai bien compris que de bonnes notes c'est loin d'être nécessaire pour une carrière en recherche. Tu m'as offert un projet de fin de baccalauréat dans ton lab qui a débouché sur une publication (tellement simpliste, mais pourtant si difficile à écrire). Pour ensuite m'offrir une place à la maîtrise ! Si un jour j'ai une carrière dans le domaine académique, c'est à toi que je la dois. Même si ça s'arrêterait demain matin, ce sont des années que je vais chérir toute ma vie.

# Abréviations

**2/3/4D** Deux/trois/quatre dimensions

**ACT** Anatomically constrained tractography

**AD** Alzheimer’s disease

**ADC** Coefficient apparent de diffusion (Apparent diffusion coefficient)

**ADNI** Alzheimer’s disease neuroimaging initiative

**AF** Faisceau arqué (Arcuate fasciculus)

**AI** Intelligence artificielle (Artificial Intelligence)

**AVC** accident vasculaire cérébral

**BA** Adjacence de faisceau (Bundle adjacency)

**BALSA** Brain Analysis Library of Spatial maps and Atlases

**BOI** Faisceau d’intérêt (Bundle of interest)

**BST** Tractographie spécifique par faisceau (Bundle-specific tractography)

**CC** Corps calleux (Corpus callosum)

**CLARITY** Clear lipid-exchanged acrylamide-hybridized rigid imaging / immunostaining / in situ-hybridization-compatible tissue hydrogel)

**CMC** Continuous map criterion

**CPU** Central processing unit

**CSD** Déconvolution sphérique contrainte (Constrained spherical deconvolution)

**CSF** Liquide cébrospinal (Cerebrospinal fluid)

**CHUS** Centre Hospitalier Universitaire de Sherbrooke

## ABRÉVIATIONS

<b>CST</b>	Faisceau corticospinal (Corticospinal tract)
<b>DTI</b>	Imagerie par tenseur de diffusion (Diffusion tensor imaging)
<b>dODF</b>	Fonction de distribution d'orientations de diffusion (Diffusion orientation distribution function)
<b>DSC</b>	Coefficient Dice (Dice score coefficient)
<b>EADC</b>	European Alzheimer's disease consortium
<b>FA</b>	Anisotropie fractionnelle (Fractional anisotropy)
<b>FAT</b>	Fronto aslant tract
<b>fODF</b>	Fonction de distribution d'orientations des fibres (Fiber orientation distribution function)
<b>FRF</b>	Fonction de réponse de fibre (Fiber response function)
<b>eFOD</b>	Enhanced fiber orientation distribution
<b>GM</b>	Substance grise (Gray matter)
<b>HARDI</b>	Imagerie de diffusion à haute résolution angulaire (High angular resolution diffusion imaging)
<b>HarP</b>	Harmonized Hippocampal Protocol
<b>HC</b>	Healthy control
<b>HCP</b>	Human Connectome Project
<b>IFOF</b>	Inferior fronto-occipital fasciculus
<b>IQR</b>	Intervalle interquartile (Interquartile range)
<b>IRM/MRI</b>	Imagerie par résonance magnétique (Magnetic resonance imaging)
<b>IRMd/dMRI</b>	Imagerie par résonance magnétique de diffusion (Diffusion magnetic resonance imaging)
<b>IRMf/fMRI</b>	Imagerie par résonance magnétique fonctionnelle (Functional magnetic resonance imaging)
<b>MD</b>	Diffusivité moyenne (Mean diffusivity)
<b>MDF</b>	Minimal direct-flip distance
<b>ML</b>	Apprentissage machine (Machine learning)

## ABRÉVIATIONS

<b>OCT</b>	Tomographie en cohérence optique (Optical coherence tomography)
<b>ODF</b>	Fonction de distribution des orientations (Orientation distribution function)
<b>OR</b>	Radiation optique (Optic radiation)
<b>PFT</b>	Particle filter tractography
<b>PLI</b>	Imagerie par lumière polarisée (Polarized light imaging)
<b>PVE</b>	Effet de volume partiel (Partial volume effect)
<b>PyT</b>	Faisceau pyramidal (Pyramidal tract)
<b>QB</b>	Quickbundles
<b>RAM</b>	Random access memory
<b>RB(x)</b>	Recobundles(X)
<b>ROI</b>	Régions d'intérêt (Regions of interest)
<b>RGB</b>	Rouge, vert et bleu (Red, green and blue)
<b>MCI</b>	Mild cognitive impairment
<b>NMR</b>	Résonance magnétique nucléaire (Nuclear magnetic resonance)
<b>SCIL</b>	Laboratoire d'Imagerie de la Connectivité de Sherbrooke (Sherbrooke connectivity imaging laboratory)
<b>SET</b>	Surface enhanced tractography
<b>SFOF</b>	Superior fronto-occipital fasciculus
<b>SLR</b>	Streamlines linear registration
<b>SH</b>	Harmoniques sphériques (Spherical harmonics)
<b>TDI</b>	Tract density imaging
<b>TODI</b>	Tract orientation density imaging
<b>UF</b>	Faisceau unciné (Uncinate fasciculus)
<b>WM</b>	Matière blanche (White matter)

“Do I really look like a guy with a  
plan?”

---

— The Joker, The Dark Knight.

# Table des matières

<b>Sommaire</b>	<b>ii</b>
<b>Remerciements</b>	<b>iv</b>
<b>Abréviations</b>	<b>vi</b>
<b>Table des matières</b>	<b>x</b>
<b>Liste des figures</b>	<b>xv</b>
<b>Liste des tableaux</b>	<b>xviii</b>
<b>Introduction</b>	<b>1</b>
<b>1 Chapitre 1 - Contexte</b>	<b>4</b>
1.1 Survol des technologies . . . . .	4
1.1.1 Le cerveau humain . . . . .	5
1.1.2 Le principe de l'IRM . . . . .	7
1.1.3 Le principe de l'IRMd . . . . .	11
1.2 Modèles locaux . . . . .	15
1.2.1 Imagerie par tenseur de diffusion . . . . .	15
1.2.2 Imagerie de diffusion à haute résolution angulaire . . . . .	18
1.3 Reconstructions par tractographie . . . . .	22
1.3.1 Tractographie locale . . . . .	23
1.3.2 Tractographie globale . . . . .	30
1.4 Méthodes de validations . . . . .	31

## TABLE DES MATIÈRES

1.4.1	Dissection post-mortem . . . . .	31
1.4.2	Coloration histologique . . . . .	33
1.4.3	Approches modernes . . . . .	36
1.5	Analyse de tractogramme . . . . .	38
1.5.1	Dissection virtuelle . . . . .	39
1.5.2	Connectome . . . . .	41
1.6	Conclusion . . . . .	42
<b>2</b>	<b>Chapitre 2 - Défis modernes</b>	<b>45</b>
2.1	Introduction . . . . .	47
2.2	Tractography versus underlying anatomy, what is and what isn't? . .	50
2.2.1	Local signal limitations . . . . .	50
2.2.2	Spatial prior limitations . . . . .	58
2.2.3	Tractogram saturation . . . . .	63
2.2.4	Obstacles along the path . . . . .	65
2.3	Individual pathways and connectomes, how to (not) interpret it? . .	68
2.3.1	Bundle segmentation . . . . .	68
2.3.2	Debate over anatomical definition of pathways . . . . .	74
2.3.3	Connectomics . . . . .	75
2.4	Discussion and Conclusion . . . . .	78
2.5	Acknowledgements . . . . .	80
<b>3</b>	<b>Chapitre 3 - Tractostorm</b>	<b>82</b>
3.1	Introduction . . . . .	86
3.1.1	Bundle segmentation . . . . .	88
3.1.2	Quantifying reproducibility in tractography . . . . .	89
3.1.3	Summary of contributions of this work . . . . .	93
3.2	Method . . . . .	94
3.2.1	Study design . . . . .	94
3.2.2	DWI datasets, processing and tractography . . . . .	95
3.2.3	Dissection plan and instructions . . . . .	98
3.2.4	Bundles analysis . . . . .	98
3.3	Results . . . . .	102



## TABLE DES MATIÈRES

3.3.1	Intra-rater evaluation . . . . .	104
3.3.2	Inter-rater evaluation . . . . .	104
3.3.3	Gold standard evaluation . . . . .	105
3.4	Discussion . . . . .	108
3.4.1	Evaluation of protocols . . . . .	108
3.4.2	Measures and representations . . . . .	109
3.4.3	Tractography algorithms . . . . .	111
3.4.4	Impact on analysis . . . . .	112
3.4.5	Future work . . . . .	113
3.5	Conclusion . . . . .	114
<b>4</b>	<b>Chapitre 4 - RecobundlesX</b>	<b>117</b>
4.1	Introduction . . . . .	117
4.1.1	Segmentation multi-atlas de faisceaux . . . . .	119
4.2	Présentation de l'algorithme Recobundles . . . . .	120
4.3	Multi-Atlas et Multi-paramètres . . . . .	122
4.3.1	Choix d'atlas . . . . .	122
4.3.2	Paramètres optimaux . . . . .	122
4.3.3	Fusion d'étiquettes . . . . .	123
4.4	Évaluation et métriques . . . . .	124
4.4.1	Dice similarity coefficient . . . . .	124
4.4.2	Bundle adjacency . . . . .	125
4.4.3	Classification binaire . . . . .	127
4.5	Expérimentations . . . . .	130
4.5.1	Jeu de données . . . . .	130
4.5.2	Ensemble de configurations . . . . .	131
4.5.3	Exécutions de l'algorithme . . . . .	131
4.6	Résultats . . . . .	132
4.6.1	Atlas basés sur les vérités terrain . . . . .	139
4.7	Utilisations externes . . . . .	142
4.7.1	Création d'atlas personnalisés . . . . .	144
4.8	Discussions . . . . .	147

## TABLE DES MATIÈRES

4.8.1	Influence du nombre d'exécutions . . . . .	147
4.8.2	Comparaison à l'approche manuelle . . . . .	149
4.9	Conclusion . . . . .	150
<b>5</b>	<b>Chapitre 5 - Tractographie spécifique par faisceau</b>	<b>152</b>
5.1	Introduction . . . . .	156
5.1.1	Whole-brain tractography . . . . .	158
5.1.2	ROI-based tractography . . . . .	159
5.2	Methods . . . . .	162
5.2.1	Building the templates of streamlines . . . . .	163
5.2.2	Anatomical priors : Tractography map creation . . . . .	166
5.2.3	Orientational priors : Enhancing fiber orientations . . . . .	166
5.2.4	DWI Datasets . . . . .	168
5.2.5	Template creation . . . . .	169
5.2.6	Experiments, Evaluation, and Statistical analysis . . . . .	170
5.3	Results . . . . .	172
5.4	Discussion . . . . .	178
5.4.1	Template creation needed for Bundle-Specific Tractography . .	179
5.4.2	Bundle volume and spatial coverage . . . . .	180
5.4.3	Reproducibility . . . . .	181
5.4.4	Computational Performance . . . . .	181
5.4.5	Future Work . . . . .	183
5.5	Conclusion . . . . .	183
	<b>Discussion</b>	<b>185</b>
	<b>Conclusion</b>	<b>197</b>
<b>A</b>	<b>Annexes et schémas</b>	<b>205</b>
A.1	Tractographie spécifique aux faisceaux . . . . .	205
A.1.1	BST, anatomical definition . . . . .	205
A.1.2	BST, template creation overview . . . . .	207
A.1.3	BST, results from all tractography algorithms . . . . .	207

## TABLE DES MATIÈRES

A.2	Bundle-specific fornix reconstruction for dual-tracer PET-tractometry	212
A.2.1	Introduction . . . . .	214
A.2.2	Methods . . . . .	216
A.2.3	Results . . . . .	219
A.2.4	Discussion . . . . .	223
A.2.5	Conclusion . . . . .	225

# Liste des figures

1.1	Schéma des principales composantes d'un axone . . . . .	6
1.2	Contraste de différents types d'acquisitions . . . . .	9
1.3	Schéma représentant le volume d'acquisition d'un IRM cérébral . . .	10
1.4	Représentation du phénomène de diffusion . . . . .	12
1.5	Diffusion et espace-q . . . . .	14
1.6	Effet de la diminution de la résolution spatiale . . . . .	15
1.7	Population de fibres et schéma de tenseur . . . . .	17
1.8	Fibercup et modèle locaux . . . . .	19
1.9	Convolution et déconvolution . . . . .	21
1.10	Champ de vecteurs et dessin anatomique . . . . .	24
1.11	Tractographie probabilistique et déterministique . . . . .	27
1.12	Schéma d'une itération de la tractographie . . . . .	29
1.13	Exemples de dissection anatomique d'un cerveau . . . . .	32
1.14	Exemples de coloration histologique . . . . .	34
1.15	Exemple d'injection de traceur axonal . . . . .	35
1.16	Exemples de méthodes de validation modernes . . . . .	37
1.17	Segmentation par régions d'intérêt du fornix . . . . .	40
1.18	Composantes à la construction d'une matrice de connectome . . . . .	42
2.1	Occurrence of the keyword tractography . . . . .	49
2.2	Bottleneck effect schematic . . . . .	52
2.3	Wall effect schematic . . . . .	53
2.4	Wall effect simulation . . . . .	54
2.5	Narrow intersection effect schematic . . . . .	55

## LISTE DES FIGURES

2.6	Three bundles with switching pathways . . . . .	57
2.7	Cortical coverage of various tractography methods . . . . .	59
2.8	Tissues classification probability across tools . . . . .	61
2.9	Visual effect of GM PVE on the IFOF . . . . .	62
2.10	Saturation curve of 3 bundles . . . . .	64
2.11	CC obstable and streamlines distribution . . . . .	67
2.12	Strict vs liberal PyT segmentation . . . . .	69
2.13	Automatic segmentation results comparison . . . . .	73
2.14	Cortical parcellation variability . . . . .	77
3.1	Illustration of the dissection plan of the PyT . . . . .	89
3.2	Representation of the Dice Coefficient . . . . .	92
3.3	Schematic of the study design . . . . .	96
3.4	Comparison of bundles and the impacts of spurious streamlines . . . .	100
3.5	Table showing the reproducibility “score” . . . . .	100
3.6	Example of average segmentation / gold standard . . . . .	102
3.7	Bar plots graph of the individual measurement . . . . .	103
3.8	Bar plots graph of the intra rater reproducibility . . . . .	104
3.9	Bar plots graph of the inter rater reproducibility . . . . .	105
3.10	Reproducibility against the gold standard (similarity measures) . . . .	106
3.11	Reproducibility against the gold standard (binary measures) . . . . .	107
4.1	Exemple de calcul de la bundle adjacency . . . . .	127
4.2	Schéma didactique des mesures de classification binaire . . . . .	130
4.3	Représentation des commandes générées pour <i>RecobundlesX</i> . . . . .	132
4.4	Comparaison visuelle des résultats de <i>Recobundles</i> et <i>RecobundlesX</i> . .	133
4.7	Mesure de classification binaire pour la représentation voxélique . . . .	138
4.8	Grille présentant plusieurs atlas de faisceaux . . . . .	143
5.1	Crossing within the centrum semiovale . . . . .	157
5.2	Challenges of the PT reconstruction . . . . .	158
5.3	Streamlines and masks template . . . . .	163
5.4	Streamlines template and priors generation . . . . .	164

## LISTE DES FIGURES

5.5	Priors and E-FOD in the centrum semiovale . . . . .	168
5.6	Single voxel FOD, prior and E-FOD . . . . .	169
5.7	Single participant qualitative results . . . . .	173
5.8	Quantitative comparison between tractography algorithms . . . . .	175
A.1	RBx pipeline overview . . . . .	207
A.2	Probabilistic particles filtering tractography (BST) . . . . .	208
A.3	Probabilistic tractography (BST) . . . . .	209
A.4	Deterministic particles filtering tractography (BST) . . . . .	210
A.5	Deterministic tractography (BST) . . . . .	211
A.6	T1 and FA template . . . . .	217
A.7	Single voxel orientation prior . . . . .	218
A.8	ROIs for fornix segmentation . . . . .	218
A.9	Average volume between methods . . . . .	220
A.10	Visual comparison between methods . . . . .	220
A.11	Single participant fornix and tract-profile . . . . .	221
A.12	Average and tract-profiles . . . . .	222

# Liste des tableaux

4.1	Résultats agrégés pour les exécutions de <i>RecobundlesX</i> . . . . .	134
4.2	Mesure de chevauchement et de distance (représentation voxélique) .	136
4.3	Mesure de classification binaire pour la représentation voxélique . . .	137
4.4	Résultats des mesures individuelles . . . . .	139
4.5	Mesure d'accord pour la représentation voxélique . . . . .	140
4.6	Mesure de classification binaire pour la représentation voxélique . . .	141
5.1	Quantitative comparison between tractography algorithms . . . . .	174

“All that is gold does not glitter, not  
all those who wander are lost”

---

— J.R.R Tolkien, *The Lord of the Rings*.



# Introduction

L'imagerie par résonance magnétique de diffusion (IRMd) est une modalité d'acquisition permettant de sonder les tissus biologiques et d'en extraire une variété d'informations sur le mouvement microscopique des molécules d'eau. Plus spécifiquement à l'imagerie médicale, l'IRMd permet l'investigation des structures fibreuses de nombreux organes et facilite la compréhension des processus cognitifs ou au diagnostic. Dans le domaine des neurosciences, l'IRMd est cruciale à l'exploration de la connectivité structurelle de la matière blanche.

Cette thèse s'intéresse plus particulièrement à la reconstruction de faisceaux de la matière blanche ainsi qu'à leur analyse. Toute la complexité du traitement des données commençant au scanneur jusqu'à la création d'un tractogramme est extrêmement importante. Par contre, l'application spécifique de reconstruction des faisceaux anatomiques plausibles est ultimement le véritable défi de l'IRMd. L'optimisation des paramètres de la tractographie, le processus de segmentation manuelle ou automatique ainsi que l'interprétation des résultats liée à ces faisceaux sont toutes des étapes du processus avec leurs lots de difficultés.

Cette thèse est organisée de la façon suivante : le Chapitre 1 introduit le concept de diffusion en physique, dans le cerveau et en imagerie médicale (sec. 1.1). Ensuite, on y présente les modèles locaux (sec. 1.2) et la tractographie (sec. 1.3) afin d'obtenir un tractogramme. Finalement, on introduit les méthodes permettant d'extraire de l'information d'un tractogramme (sec. 1.5).

## INTRODUCTION

Une fois les bases fondamentales de l'IRMd et de la tractographie bien établies, il est crucial de bien comprendre les limitations de ce domaine. Le Chapitre 2 présente un aperçu général des différents défis pour notre domaine ainsi que plusieurs perceptions erronées, biais et pièges de la tractographie. Cette présentation didactique permet d'avoir un résumé rapide des limitations de la tractographie et d'avoir une meilleure perspective du domaine.

Le Chapitre 3 analyse en détail la variabilité expérimentale liée à la dissection virtuelle de faisceaux. À l'aide d'une conception méthodologique visant une évaluation à l'aveugle, une vingtaine de participants ont suivi des instructions permettant d'effectuer une dissection virtuelle. Grâce à cette méthodologie la variabilité intra et inter expert peut être évaluée. Cette quantification de la variabilité humaine permettra de mieux évaluer les performances d'algorithmes de segmentation automatiques.

Grâce à la quantification de la variabilité et la création d'un jeu de données de segmentations manuelles expertes provenant du projet Tractostorm, il est désormais possible de mieux évaluer des algorithmes automatiques. Le Chapitre 4 présente un algorithme de segmentation automatique reposant sur une approche multi-atlas et multi-paramètres suivi d'une fusion d'étiquettes. La Section 4.1.1 explique l'algorithme sous-jacent (Recobundles), tandis que la Section 4.3 présente la contribution liée à l'approche multi-atlas et multi-paramètres. Finalement, la Section 4.4 explique comment les résultats peuvent être évalués quantitativement à l'aide d'une variété de mesures.

Le Chapitre 5 décrit une nouvelle approche de tractographie par faisceau permettant d'améliorer la reconstruction grâce à des *a priori* anatomiques. À l'aide de modèles de faisceaux basés sur des connaissances anatomiques, des *a priori* spatiaux (sec. 5.2.2) et d'orientations locales (sec. 5.2.3) peuvent être pré-calculés et injectés dans les algorithmes de tractographie. Cette approche permet d'optimiser le traitement de données pour spécialiser la tractographie lorsqu'une hypothèse ne concerne que quelques faisceaux plutôt que l'ensemble de la connectivité structurelle du cerveau.

“Somewhere, something incredible is  
waiting to be known.”

---

— Carl Sagan, *Cosmos*.

# Chapitre 1

## Contexte

### 1.1 Survol des technologies

Le fondement de l'imagerie par résonance magnétique de diffusion (IRMd) vient de la résonance magnétique nucléaire (RMN). En 1946 Bloch [1946] et Purcell et al. [1946] ont observé que certains noyaux atomiques émettaient un signal lorsqu'ils étaient introduits dans un champ magnétique. Pour faire le saut entre signal RMN et acquisition IRM, il a fallu attendre l'introduction de la localisation spatiale par Lauterbur [1973] et le développement de la séquence d'imagerie écho planaire (EPI) par Mansfield [1977]. Cet avancement permet d'étudier la composition de tissus biologiques de façon non-invasive. Le signal émis par les protons d'hydrogène présents dans les tissus cérébraux permet de reconstruire une image où, par exemple, la matière blanche, la matière grise et le liquide cébrospinal (LCS) vont avoir un intervalle de contraste permettant de les différencier.

Introduite en 1985 par Le Bihan et Breton [1985], l'IRMd permet, quant à elle, d'observer les mouvements microscopiques des molécules d'eau. À l'origine, ce contraste était utilisé pour l'estimation des zones touchées par des accidents vasculaires cérébraux dans un contexte d'application clinique [Moseley et al., 1990; Van Everdingen et al., 1998; Warach et al., 1992]. Basser et al. [1994] introduit par la suite l'imagerie par tenseur de diffusion (DTI) qui permet d'inférer l'orientation des structures

## 1.1. SURVOL DES TECHNOLOGIES

de la matière blanche. Le développement de la tractographie, basée sur le tenseur de diffusion, a rapidement suivi. La tractographie a permis l'estimation à une plus grande échelle de la connectivité cérébrale. Cet effort collectif de «*cartographier*» les connexions du cerveau humain se nomme *Connectome* [Hagmann et al., 2008].

### 1.1.1 Le cerveau humain

À l'intérieur du crâne, dans moins de  $1500\text{cm}^3$ , sont entassés entre 80 et 90 milliards de neurones formant plus de 100 billions (100 mille milliards) de connexions. De façon très simplifiée, le cerveau est généralement décrit comme ayant trois principaux milieux ; la matière grise, la matière blanche et le liquide cébrospinal. Le cerveau est composé d'une variété de cellules tout utiles à son bon fonctionnement, dans le cadre de cette thèse le type de cellule d'intérêt est le neurone. Une vue d'ensemble du cerveau, de certaines régions d'intérêt ainsi que de l'anatomie d'un neurone sont illustrées à la Figure 1.1. La matière grise est composée du corps cellulaire des neurones et se retrouve à la surface du cerveau, le cortex, et dans les noyaux centraux. La matière blanche, quant à elle, est formée des axones et agit en tant que circuiterie responsable du transport de l'information. Ces milliards d'axones densément organisés sont typiquement entourés d'une gaine de myéline, ce qui donne à la matière blanche sa coloration plus pâle et sa texture fibreuse. Le liquide cébrospinal permet non seulement d'agir en tant que coussin protecteur contre les chocs, mais permet aussi de maintenir une pression stable à l'intérieur du crâne ainsi qu'au transport de certains nutriments et à l'élimination de déchets.

La longueur totale des axones du cerveau dépasserait les centaines de milliers de kilomètres s'ils étaient mis bout à bout. Étant contrainte à l'intérieur du crâne, l'organisation des axones et de leurs connexions n'est pas aléatoirement distribuée. Les axones ayant des origines et des terminaisons spatialement cohérentes se regroupent entre elles, sous forme de faisceaux (fasciculation), pour parcourir la matière blanche. Ces faisceaux peuvent être séparés en 3 grandes catégories ; les faisceaux d'associations, commissuraux et de projections. Les faisceaux d'associations assurent la connexion entre les différentes régions corticales d'un même hémisphère. Les faisceaux commissuraux connectent des zones corticales inter-hémisphériques (axe

## 1.1. SURVOL DES TECHNOLOGIES

gauche-droit). Finalement, les faisceaux de projections assurent le transport de l'information entre le cortex et les noyaux gris centraux et la moelle épinière (axe haut-bas).

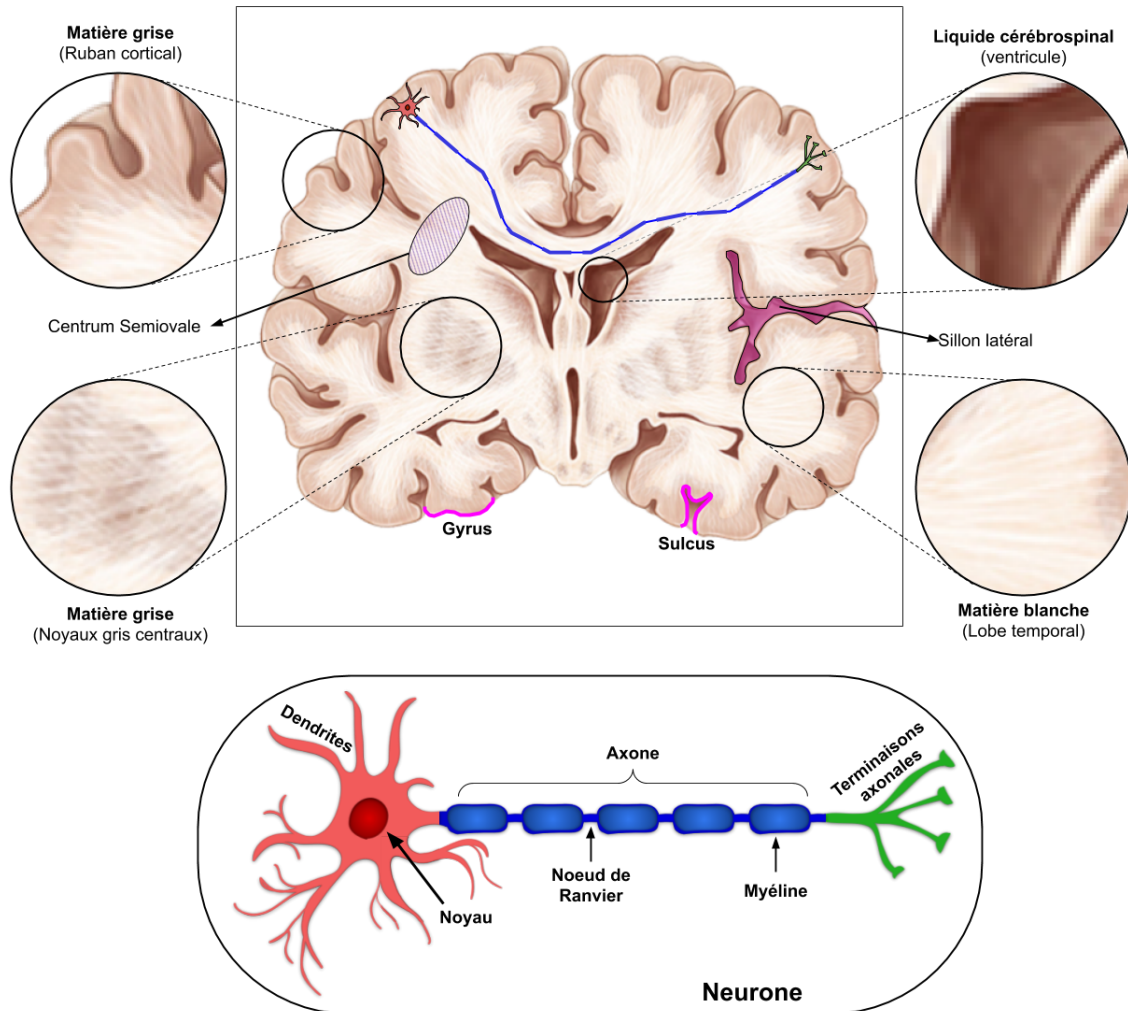


figure 1.1 – Schéma des principales composantes d'un axone et de structures d'intérêts dans le cerveau. Comme illustré, les corps cellulaires et noyaux ainsi que les terminaisons axonales des neurones ne se retrouvent que dans la matière grise. La seule composante d'un neurone se retrouvant dans la matière blanche est l'axone. Les circonvolutions (gyri et sulci) du cerveau possèdent leurs propres nomenclatures et sont généralement des points de référence pour l'analyse de la connectivité structurelle. La région identifiée comme *Centrum semiovale* est particulièrement intéressante due au fait que des populations de fibres d'associations, commissurales et de projections s'y croisent.

## 1.1. SURVOL DES TECHNOLOGIES

Divers faisceaux et les zones corticales qu'ils connectent sont associés à des fonctions précises. Par exemple, certaines régions du lobe occipital sont liées au traitement de l'information visuelle. Cette région est connectée au thalamus (noyau géniculé latéral) par les radiations optiques et leur permet de “*communiquer*” (voir Figure 1.13C). Le faisceau arqué relie l'aire de Broca, dans le lobe frontal inférieur, et l'aire de Wernicke, dans le lobe temporal supérieur (voir Figure 1.13B). Ce faisceau et les deux principales régions qu'il connecte sont fortement liés aux théories du langage et sont, dans un contexte clinique, liés à l'aphasie. Un dernier exemple est le faisceau corticospinal reliant, entre autres, les régions motrices adjacentes au sulcus central et la moelle épinière (voir Figure 1.13D). Ces régions corticales permettent la transmission d'informations impliquant la planification, le contrôle et l'exécution des mouvements volontaires des muscles du corps.

Ces faisceaux possèdent une morphologie connue, relativement similaire d'un individu à l'autre et contribuent à des fonctions similaires. Beaucoup d'inconnus sont encore présents en ce qui concerne, entre autres, l'évolution de ces faisceaux au cours du développement/vieillessement, l'effet spécifique de certaines pathologies sur la morphologie ou la qualité de l'information transmise par ces faisceaux.

### 1.1.2 Le principe de l'IRM

Le processus physique fondamental à l'IRM repose sur le fait que le spin des atomes s'aligne avec la directionnalité du champ magnétique dans lequel ils sont plongés. Le type d'atome d'intérêt pour l'IRM est l'hydrogène, ces atomes d'hydrogène sont présents en grande quantité dans le corps humain dû à sa haute composition en eau (environ 70%). En présence d'un fort champ magnétique uniforme, ces spins vont être orientés de la même façon et avoir la même fréquence de précession (effectuer une révolution complète sur son axe de rotation). Ceci est dû au fait que l'atome d'hydrogène n'est qu'un proton et celui-ci se comporte comme un petit aimant. Cette fréquence, nommée fréquence de Larmor, dépend de la force du champ magnétique. À l'aide d'ondes électromagnétiques, ces révolutions sont synchronisées, ou en phase ; le signal de l'IRM émerge de ce phénomène de résonance. Sans ces ondes électromagnétiques, les spins resteront alignés avec le champ magnétique, mais perdront plus

## 1.1. SURVOL DES TECHNOLOGIES

ou moins rapidement leur synchronisation et donc leur signal de résonance.

La vitesse à laquelle ces spins vont retourner à leur état initial va dépendre des molécules auxquelles sont attachés les atomes d'hydrogène. C'est pourquoi chaque tissu possède des propriétés physiques différentes et celles-ci influencent le signal observé. La différence dans l'intensité observée du signal entre les tissus est nommée *contraste*. Un scanner IRM peut être programmé de façon à observer différemment une ou plusieurs propriétés physiques plutôt que d'autres. Ces différents types d'acquisitions sont plus sensibles à certains tissus. Ces programmations au scanner se nomment *séquences*. La Figure 1.2 présente plusieurs contrastes d'un même cerveau acquis à l'aide de différentes séquences au scanner IRM.



## 1.1. SURVOL DES TECHNOLOGIES

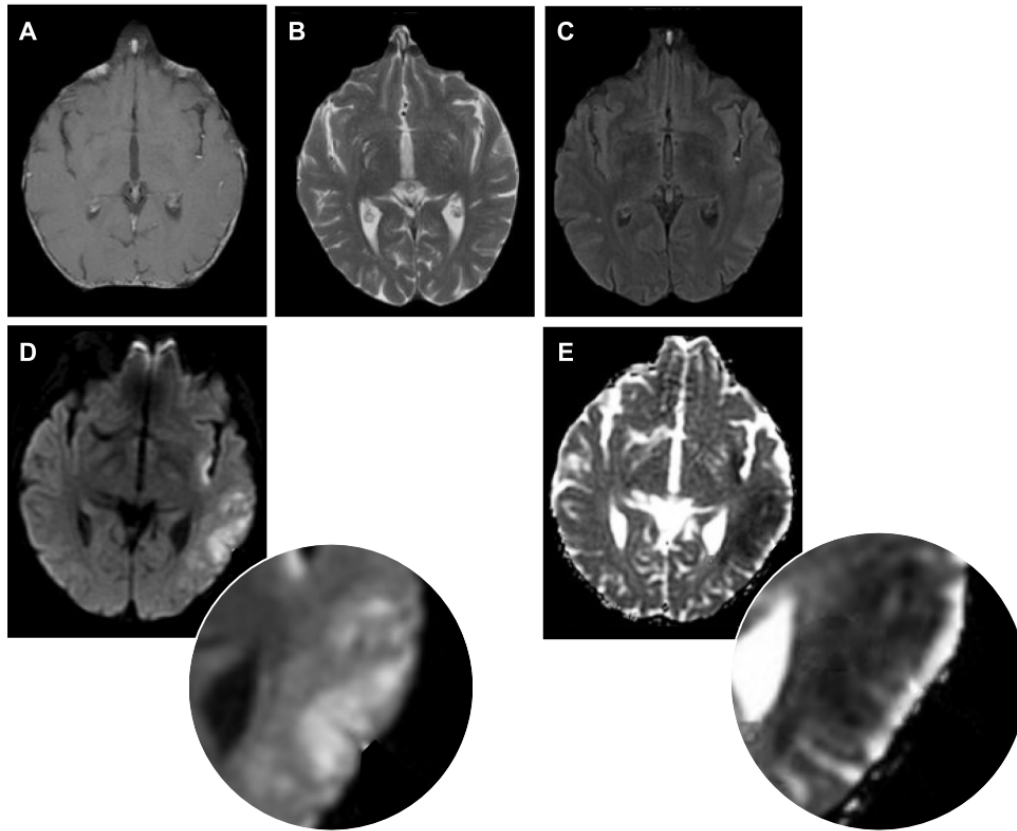


figure 1.2 – Contraste de différents types d’acquisitions permettant de sonder différentes propriétés physiques et ainsi observer, par exemple, des tissus anormaux ou pathologiques. La présence d’un infarctus ischémique aigu est impossible à identifier sur l’acquisition de type T1 (A), T2 (B) et FLAIR (C), mais facilement identifiable par la zone d’hyper-intensité en imagerie par résonance magnétique de diffusion (D) ou en hypo-intensité au volume représentant le coefficient de diffusion apparent (E). En vignette, la zone d’intérêt est mise en évidence. Image adaptée de [Roldan-Valadez et Lopez-Mejia, 2014].

Le résultat d’une acquisition IRM est typiquement un volume, ou grille, en 3 dimensions (3D). Cette grille est composée d’éléments de volume (voxel) et la valeur associée est proportionnelle à l’intensité du signal généré par les spins en résonance de ce voxel. Ces volumes peuvent être visualisés par tranche de façon à faciliter l’interprétation des données (contraste entre les tissus, morphologies, point de repère anatomique, etc.). La Figure 1.3 illustre cette grille ainsi que la terminologie associée

## 1.1. SURVOL DES TECHNOLOGIES

aux différentes coupes et orientations utilisées en imagerie médicale.

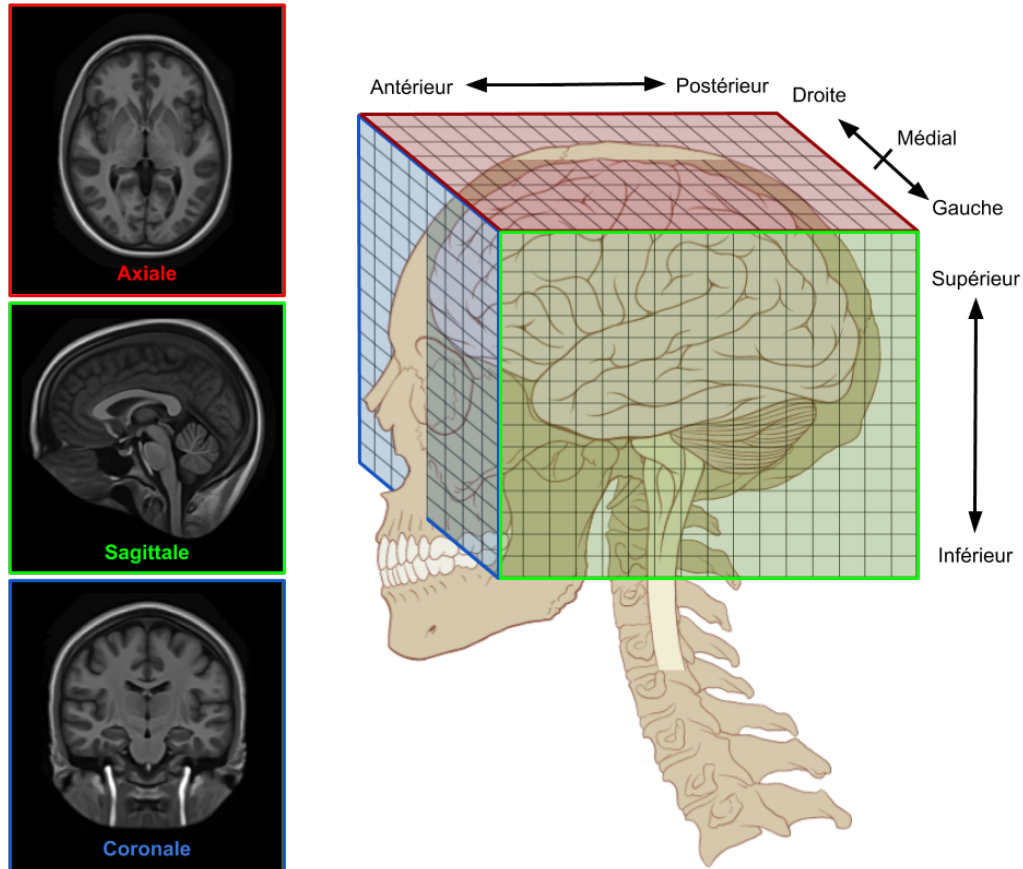


figure 1.3 – Schéma représentant le volume d'acquisition (dans le cas d'une IRM cérébrale). Les différentes coupes perpendiculaires utiles à la visualisation sont identifiées et colorées. Plusieurs nomenclatures existent pour définir la position et l'orientation des structures dans le cerveau, mais non nécessaires dans le cadre de cette thèse.

## 1.1. SURVOL DES TECHNOLOGIES

### 1.1.3 Le principe de l'IRMd

L'IRMd repose sur le même phénomène de résonance, mais la mesure du signal n'est pas prise immédiatement au moment où les spins sont en résonance (en phase). Un temps d'attente est plutôt ajouté pour permettre un retour à l'état original (déphasage), et ensuite une nouvelle onde électromagnétique inverse ce processus dans le but de retourner à l'état de résonance (rephasage). Sans diffusion, ou si la diffusion était uniforme dans toutes les directions (i.e diffusion libre), le résultat après la seconde onde électromagnétique serait un retour identique à l'état de phase initiale. De cette façon, il est possible de détecter si un déplacement des atomes d'hydrogène s'est effectué.

Toutefois, le cerveau contient différents tissus et liquides, chacun ayant leur propre caractéristique de diffusion. Le liquide cébrospinal est un exemple de diffusion libre (ou isotrope). Les molécules d'eau bougent librement et uniformément dans toutes les directions, avec aucune dépendance directionnelle. À l'opposé de cette absence complète de structure, la matière blanche est composée d'une structure fibreuse organisée contraignant le mouvement des molécules d'eau. Ce milieu est un exemple de diffusion restreinte (ou anisotrope). Les molécules d'eau bougent plus difficilement dans certaines directions, avec une forte dépendance directionnelle. Cette notion de restriction variant selon l'orientation est à la base de la capacité à inférer l'architecture sous-jacente de la matière blanche. La Figure 1.4 montre comment une structure, ou son absence, peut restreindre ou faciliter le phénomène de diffusion selon l'orientation. L'expression *population de fibres* réfère au fait qu'un groupe d'axones est localement orienté de la même façon. Lorsque deux groupes d'axones, chacun ayant son orientation, se croisent à un certain endroit, cette position est décrite comme ayant *deux populations de fibres*.

## 1.1. SURVOL DES TECHNOLOGIES

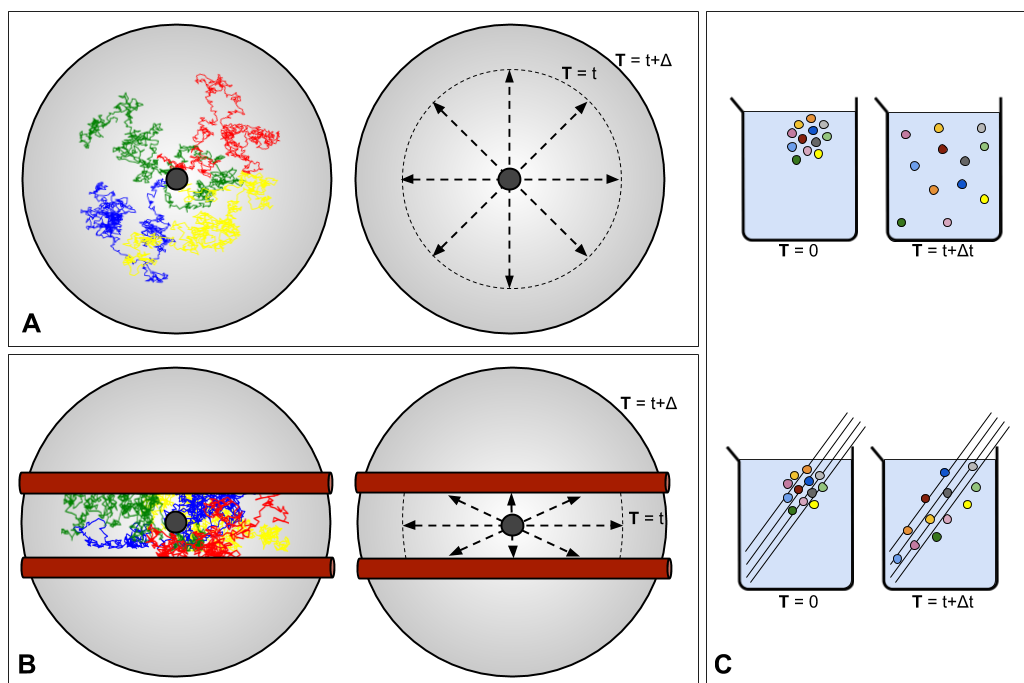


figure 1.4 – Schéma représentant le phénomène de diffusion à l'échelle microscopique et macroscopique. En A, une particule dans un milieu sans restriction (i.e diffusion isotrope). À gauche, les trajectoires colorées représentent des possibilités autant probables l'une que l'autre. À droite, la ligne pointillée représente la distance maximale pouvant être atteinte en un temps  $T$ . La probabilité d'atteindre une position décroît à mesure que la distance du centre augmente. En B), la particule est plutôt dans un milieu restreint (i.e diffusion anisotrope). À gauche, les trajectoires colorées représentent des chemins maintenant restreints par un obstacle. À droite, la probabilité d'atteindre une certaine position est maintenant affectée par l'obstacle. En C, à l'échelle macroscopique une restriction physique peut limiter la propagation d'un groupe de molécules. Par exemple, un colorant ajouté à un contenant d'eau va se propager uniformément dans toutes les directions, alors que la présence d'obstacles ralentirait la propagation au travers de ceux-ci.

Le contraste de l'IMRd émerge de ces différences entre milieu libre et milieu restreint. Le mouvement des molécules d'eau va empêcher un retour parfait à l'état initial et donc réduire le signal. Ce contraste permet de visualiser la facilité de diffusion des molécules d'eau, ou plutôt visualiser le niveau de restriction. Une des premières applications cliniques de l'IRMd était liée aux accidents vasculaires cérébraux (AVC) et utilisait, par exemple, le contraste de coefficient de diffusion apparent (*average diffu-*

## 1.1. SURVOL DES TECHNOLOGIES

*sion coefficient*, ADC). Dans le cas d'un accident vasculaire cérébral, le saignement à l'intérieur du cerveau va affecter l'intégrité des tissus touchés, l'ADC sera donc différent de celui attendu dans un tissu sain. Par contre, ceci sera invisible sur d'autres types de contraste. Les contrastes d'IMRd ont été et sont toujours de précieux outils pour le diagnostic précoce d'AVC. Tel qu'illustré à la Figure 1.2, au cours des premières heures d'un infarctus ischémique aigu il est très difficile d'identifier une lésion sur les contrastes T1, T2 ou FLAIR, mais facilement identifiable sur les contrastes liés à la diffusion.

Au-delà de la différenciation entre milieu libre ou milieu restreint, l'IRMd permet d'observer le phénomène de diffusion dans une direction particulière. Il est possible d'encoder la position spatiale au moment où les spins sont mis en phase. La diffusion se produit, les spins se déphasent et à la tentative de rephasage une différence entre l'état initial et l'état courant est observable. Cette différence est indicatrice de l'amplitude du mouvement effectué et de la direction de celui-ci. Cependant, cet encodage de la position spatiale ne peut se faire que par rapport à une direction en particulier. Il est donc possible de répéter cette observation dans plusieurs orientations de façon subséquente. Ce type d'acquisition produit donc un volume, ou une grille, en 4 dimensions (4D) où la dernière dimension représente ces différentes directions (voir Figure 1.5). La manière de résumer ces informations est présentée en plus de détails dans la Section 1.2.

## 1.1. SURVOL DES TECHNOLOGIES

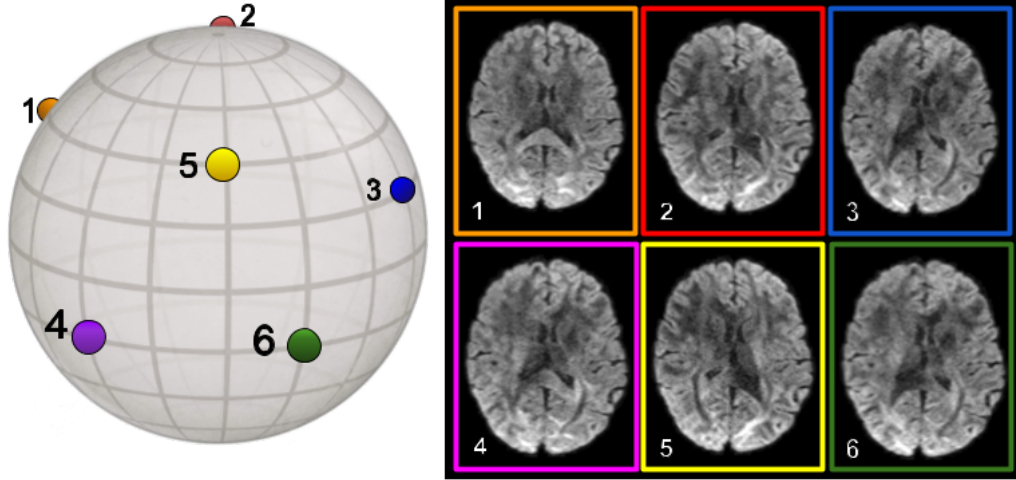


figure 1.5 – Le phénomène de diffusion est sondé en IMRd en observant celui-ci dans une direction à la fois. Le résultat est un volume (grille) 3D présentant des intensités à chaque voxel liées à l’amplitude de la diffusion dans la direction observée. Le choix de direction dans lequel observer le phénomène de diffusion consiste à échantillonner un espace nommé *espace- $q$* , ou  *$q$ -space*. La sphère, à gauche, montre 6 directions d’échantillonnage dans l’*espace- $q$*  et les coupes axiales, à droite, montre les intensités observées à chacune de ces directions. Dans le cas du tenseur de diffusion, le nombre minimal de directions à échantillonner pour obtenir un ensemble d’observations adéquat et suffisant pour le modèle est de 6 (12 pour préserver l’invariabilité rotationnelle). Ces 6 (ou 12) directions doivent être uniformément distribuées sur une sphère pour bien modéliser le processus de diffusion. Pour les algorithmes plus complexes liés à l’imagerie de diffusion à haute résolution angulaire, cet échantillonnage peut atteindre plus de 100 directions et rapidement se complexifier.

Cette grille, composée de voxels, peut avoir une basse ou haute résolution. La Figure 1.6 montre les effets d’une perte de résolution spatiale. Typiquement, en IRMd une haute résolution est de l’ordre de 1-1.5mm, ce qui veut dire que chaque voxel représente un volume entre  $1mm^3$  et  $3.375mm^3$ . À ces résolutions, le cerveau entier peut être représenté dans une grille d’environ  $128 \times 128 \times 128$  voxels. Tandis qu’une image IRMd de basse résolution est plutôt de l’ordre de 2mm-2.5mm et donc chaque voxel a un volume entre  $8mm^3$  et  $15.625mm^3$ . Avec ces voxels de plus grande taille, le cerveau peut être représenté à l’aide d’une grille d’environ seulement  $64 \times 64 \times 64$  voxels.

## 1.2. MODÈLES LOCAUX

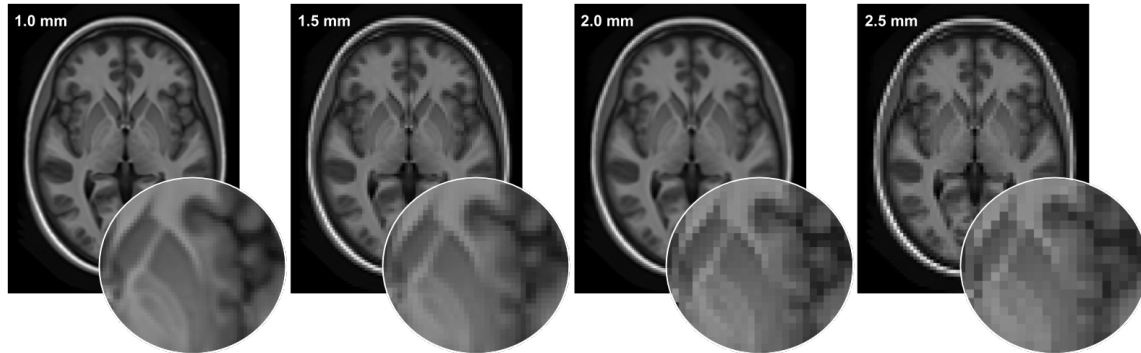


figure 1.6 – L’effet de la diminution de la résolution spatiale sur les détails (en vignette) d’une image pondérée en T1. La résolution en millimètres indique la taille de chaque élément de la grille (voxel) et toutes structures de tailles inférieures à la résolution seront impossibles à distinguer.

## 1.2 Modèles locaux

Les acquisitions d’IRMd sont difficilement interprétables sous la forme de volume 4D, une étape de modélisation est donc nécessaire pour représenter la directionnalité de la structure sous-jacente. L’information de la 4e dimension est alors résumée par un *modèle local* qui est typiquement représenté par un glyphe à chaque voxel du volume 3D. Ces glyphes permettent de facilement interpréter l’orientation générale des structures au sein d’un voxel.

### 1.2.1 Imagerie par tenseur de diffusion

L’imagerie par tenseur de diffusion (DTI) [[Basser et al., 1994](#)] a été proposée pour estimer l’anisotropie et l’orientation de la diffusion de l’eau. Le modèle, le tenseur de diffusion, repose sur une hypothèse que le phénomène de diffusion est gaussien, ce qui implique que chaque voxel ne peut posséder qu’une seule population de fibres. Il est estimé que plus de 70% du volume de la matière blanche possèdent plus d’une population de fibres. Malgré cette limitation, le DTI a été, et est probablement toujours, le modèle le plus courant pour modéliser la matière blanche. La simplicité du modèle et le court temps d’acquisition au scanneur ont certainement participé à la popularité du DTI. En effet, il suffit d’observer la diffusion d’au moins 6 orientations

## 1.2. MODÈLES LOCAUX

pour obtenir une approximation adéquate du tenseur de diffusion (voir Figure 1.7). Les faisceaux majeurs du cerveau peuvent tout de même être reconstruits avec une certaine concordance par rapport à l'anatomie connue du cerveau (voir Section 1.3).

Les propriétés mathématiques du tenseur de diffusion permettent la création de cartes 3D résumant le comportement général de la diffusion dans le cerveau. Un tenseur peut être visualisé comme un glyphe (i.e un ellipsoïde 3D), cet ellipsoïde possède 3 composantes principales ( $\lambda_1, \lambda_2, \lambda_3$ ), tel que présenté à la Figure 1.7, qui le résume parfaitement. Dans une situation où le milieu présente une diffusion isotrope (égale dans toute les directions), le tenseur sera plutôt sphérique et  $\lambda_1, \lambda_2$  et  $\lambda_3$  seront égaux. La mesure la plus connue et utilisée du DTI est la carte d'anisotropie fractionnelle (FA) (voir Formule 1.1) qui représente le degré d'anisotropie dans chaque voxel [Pierpaoli et Basser, 1996]. Un voxel ayant une valeur de 1 représente un milieu parfaitement anisotrope (diffusion dans une seule direction et aucune autre), tandis qu'une valeur de FA de 0 est censée signifier un milieu parfaitement isotrope. Par contre, dû à l'hypothèse de diffusion gaussienne du DTI, un voxel contenant un croisement à 3 directions orthogonales produirait aussi une FA ayant une valeur nulle. Ceci est la principale limitation du DTI : ce modèle local n'est pas suffisamment spécifique, car plusieurs configurations peuvent faire émerger la même mesure. D'autres cartes telles que la diffusion moyenne (MD, 1.2), la diffusion radiale (RD, 1.3) ou la diffusion axiale (1.4) proviennent également du tenseur. Ces mesures sont calculées à l'aide des principales composantes du tenseur (vecteurs propres).

$$FA = \sqrt{\frac{1}{2}} \cdot \sqrt{\frac{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \quad (1.1)$$

$$MD = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \quad (1.2)$$

$$RD = \frac{\lambda_2 + \lambda_3}{2} \quad (1.3)$$

$$AD = \lambda_1 \quad (1.4)$$



## 1.2. MODÈLES LOCAUX

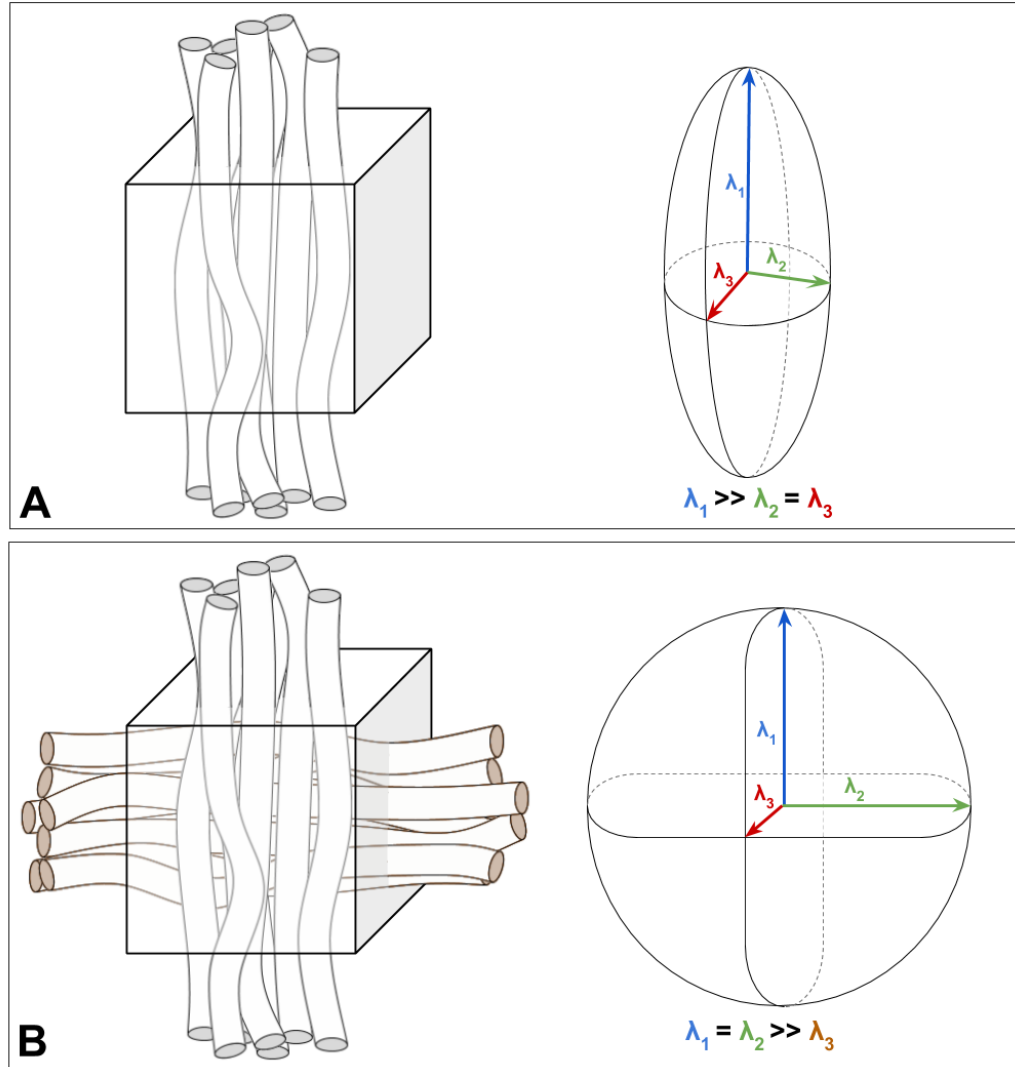


figure 1.7 – Schéma représentant la relation entre une population de fibres et le tenseur qui la représentera. En A, un voxel contenant une seule population de fibres sera facilement représentable par un tenseur. Le résultat est un ellipsoïde 3D orienté dans le sens de la population de fibres à ce voxel ( $\lambda_1$ ). Cette représentation est particulièrement intuitive à l'échelle d'un voxel avec une seule population de fibres. En B, deux populations de fibres se croisent perpendiculairement à l'intérieur du voxel, donc l'hypothèse que le processus de diffusion est gaussien n'est pas respectée. La conséquence sur le tenseur est une augmentation de la seconde composante du tenseur ( $\lambda_2$ ) et donc un ellipsoïde ayant une moins bonne correspondance aux populations de fibres sous-jacentes. L'extension à un triple croisement perpendiculaire produira un tenseur sphérique, indifférenciable d'un milieu où la diffusion est isotrope.

## 1.2. MODÈLES LOCAUX

### 1.2.2 Imagerie de diffusion à haute résolution angulaire

Comparativement au DTI, l'imagerie de diffusion à haute résolution angulaire (HARDI) permet de mieux modéliser les tissus sous-jacents [Descoteaux, 2015; Tuch et al., 2002; Wedeen et al., 2008]. L'expression DTI lie ensemble le type d'acquisition et le modèle de tenseur de diffusion. Par contre, l'expression HARDI fait uniquement référence au type d'acquisition et plusieurs types de modélisations sont possibles. Ces modèles sont plus complexes et ne reposent pas sur l'hypothèse que le phénomène de diffusion est un processus gaussien. Les voxels contenant plus d'une population de fibres peuvent donc être modélisés, en théorie, plus fidèlement. Surmonter cette limitation vient au coût d'un temps d'acquisition plus long et d'un traitement de données plus complexe.

Ces modélisations nécessitent plus d'observation du phénomène de diffusion de l'eau. Les protocoles d'acquisition HARDI varient en complexité, certains nécessitent un nombre d'observations de l'ordre de 60 à 150 directions uniformément distribuées sur la sphère (voir Figure 1.5). Tandis que l'imagerie par spectre de diffusion (DSI) par exemple [Wedeen et al., 2008] requiert un échantillonnage sous forme de grille allant jusqu'à plus de 500 directions. La Figure 1.8 présente 3 modèles locaux dans une région de croisement, où deux de ces modèles sont présentés ci-dessous et le troisième est le tenseur de diffusion présenté à la section précédente. Les mentions liées aux acquisitions IRMd présentes dans cette thèse font référence aux acquisitions de type HARDI (entre 60 et 100 directions).

## 1.2. MODÈLES LOCAUX

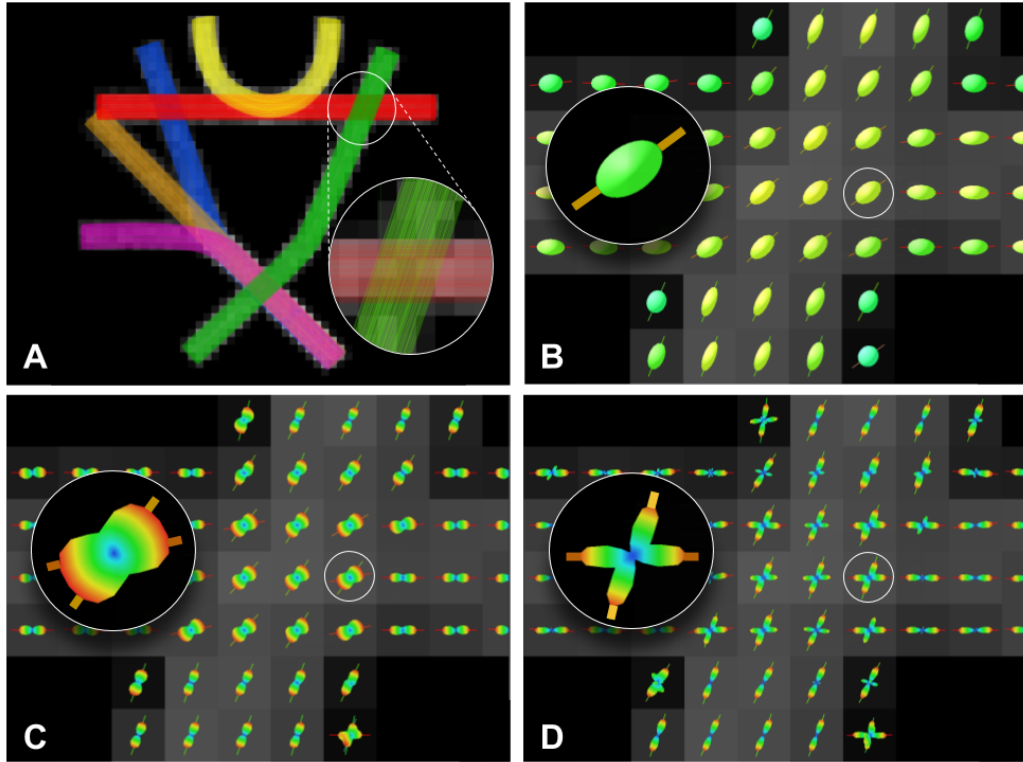


figure 1.8 – Présentation de différents modèles locaux. En A, une architecture simplifiée représentant des configurations de croisements, courbures et chevauchements afin de faciliter l'évaluation d'algorithmes liés à l'IRMd. En B, un croisement (agrandi de A) montrant un champ de tenseurs. Généralement bien aligné avec les populations de fibres hors croisement (rouge et vert en A), mais sévèrement limité dans le croisement. On peut observer l'axe principal du tenseur représenté par un vecteur (ligne 3D). En C, un champ de dODF permettant la représentation de configuration plus complexe, tel un croisement. En vignette, une seule dODF où la coloration rouge représente une forte probabilité de diffusion de l'eau, et donc possiblement la présence de population de fibres. On peut y voir 2 vecteurs, représentant chacun un maxima local de probabilité. En D, un champ de fODF permettant de représenter des croisements à angle plus aigu qu'un champ de dODF. En vignette, une seule fODF où la coloration rouge indique une forte probabilité de présence de population de fibres. Les deux maxima locaux (vecteurs) sont plus faciles à distinguer et leurs angles plus cohérents avec leurs voisins qu'en C.

## 1.2. MODÈLES LOCAUX

### Multi-tenseurs

Une extension du tenseur de diffusion, permettant de surmonter les difficultés liées au croisement de plusieurs populations de fibres, consiste à simplement modéliser plus d'un tenseur par voxel [Kreher et al., 2005; Peled et al., 2006; Tuch et al., 2002]. L'objectif étant de créer des *compartiments*, i.e tenseurs, permettant de modéliser le signal le plus fidèlement possible en ajustant les paramètres de ces tenseurs. Un défi de cette méthode est de déterminer le nombre optimal de tenseurs devant être présents dans un voxel. Par exemple, forcer deux tenseurs dans un voxel où une seule population de fibres est présente produira des résultats instables et généralement moins valides. Plusieurs approches ont été proposées pour mieux estimer ces hyper-paramètres; régulariser spatialement les estimations ou ajouter des compartiments autres que les tenseurs pour faciliter la modélisation. Une revue de ces approches a récemment été publiée par Filatova et al. [2018].

### Fonction de distribution des orientations de diffusion (dODF)

La fonction de distribution des orientations de diffusion permet de représenter la diffusion de l'eau dans chaque voxel [Aganj et al., 2010; Descoteaux et al., 2007; Tuch, 2004]. Une dODF est une distribution de probabilité sur une sphère. Pour chaque direction, l'intensité associée indique la probabilité de diffusion de l'eau dans cette direction. Le rayon de cette sphère peut être pondéré par l'intensité et ainsi créer un glyphe (voir Figure 1.8C) qui indique l'orientation générale du processus de diffusion dans ce voxel. Cette représentation permet de facilement montrer la distribution de probabilité des directions principales de diffusion. Les maxima locaux sur la sphère indiquant la principale direction de diffusion et, indirectement, la présence probable d'une population de fibres dans la même orientation. Cette représentation permet, indirectement, de modéliser des configurations de fibres plus complexes qu'un simple tenseur par exemple. Par contre, une dODF est limitée par le fait que la distribution autour des maxima locaux est très évasée (aplatie) [Descoteaux et al., 2007, 2009]. À mesure que l'angle de croisement des fibres sous-jacentes se réduit, la capacité à correctement estimer l'orientation des populations de fibres décroît rapidement.

## 1.2. MODÈLES LOCAUX

### Fonction de distribution des orientations de fibres (fODF)

Une solution pour faciliter la distinction des populations de fibres est d'*aiguiser* la dODF et ainsi mieux modéliser les configurations de croisement de fibres, même à des angles inférieurs [Descoteaux et al., 2007; Tournier et al., 2007]. La manière la plus intuitive d'expliquer le calcul des fODF est l'opération de déconvolution à partir des dODF. L'opération de déconvolution telle que présentée par [Descoteaux et al., 2009] s'effectue en utilisant une fonction de réponse de fibres et de nombreux outils mathématiques, e.g harmoniques sphériques (SH), la Figure 1.9 présente l'idée générale de l'opération de déconvolution.

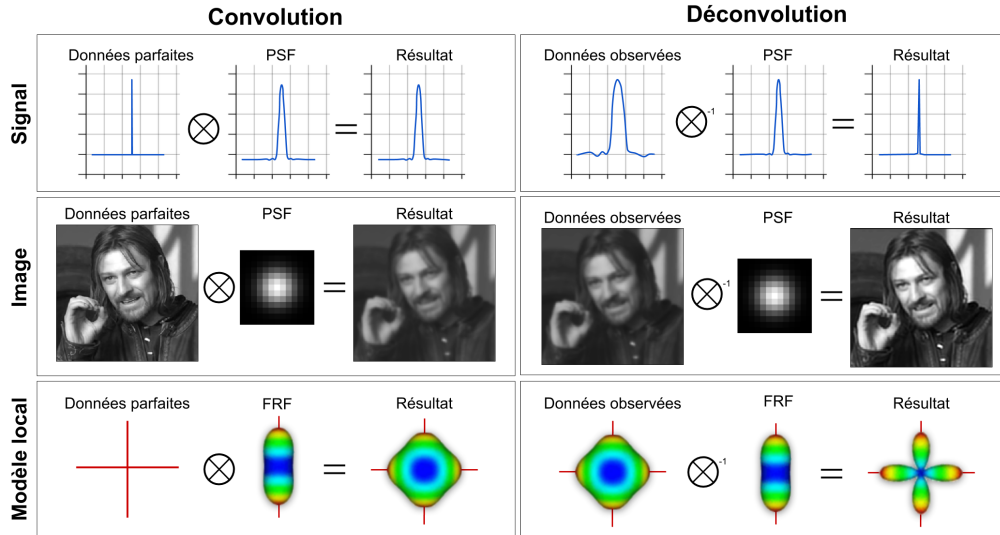


figure 1.9 – L'opération de convolution consiste à modifier un signal à l'aide d'une fonction d'étalement (PSF) et peut être comparée à ajouter un flou. L'opération de déconvolution est l'opération inverse. Considérant des données imparfaites et une fonction d'étalement connue, il est possible d'approximer les données originales. Dans le cas de la photographie 2D, la fonction d'étalement est souvent fournie par le fabricant, la déconvolution peut être utilisée pour corriger des défauts de lentilles ou de capteurs et ainsi produire des images plus fidèles à la réalité. Dans le cas spécifique des dODF et fODF, l'équivalent de la fonction d'étalement est la fonction de réponse de fibre (FRF). Celle-ci représente le signal observé dans une région possédant une seule population de fibres (e.g corps calleux). Lorsqu'un signal provenant d'une région a un nombre inconnu de populations de fibres et possède une dODF «*floue*», cette dODF est déconvoluée avec la FRF et on obtient ainsi la fODF, plus aiguisée.

### 1.3. RECONSTRUCTIONS PAR TRACTOGRAPHIE

Depuis cette contribution, plusieurs méthodes ont été proposées pour améliorer et stabiliser les résultats du calcul de fODF. En effet, l'opération de déconvolution est parfois sensible au bruit ou aux artéfacts d'oscillation. Pour n'en nommer que quelques-unes de ces améliorations ; régularisations [Descoteaux et al., 2007; Sakaie et Lowe, 2007], estimation itérative [Tax et al., 2014], fonction de réponse de fibre par tissu [Jeurissen et al., 2014] ou de meilleures méthodes de calcul de solution des systèmes d'équations [Dell'Acqua et al., 2010]. Toute la complexité des nombreux algorithmes de déconvolution sphérique est présentée dans de récentes revues de littérature [Canales-Rodríguez et al., 2019; Dell'Acqua et Tournier, 2019]. Les mentions liées au fODF présentes dans cette thèse font référence à la représentation proposée par [Descoteaux et al., 2007].

## 1.3 Reconstructions par tractographie

Le *Connectome*, ou la cartographie des connexions du cerveau, est un défi de taille qui suit les étapes déjà très complexes précédemment mentionnées. Une fois que l'acquisition IRMd est traitée, qu'un modèle local a été choisi et calculé, l'étape suivante consiste à reconstruire les chemins probables représentant la connectivité sous-jacente du cerveau [Catani et al., 2002; Jones et al., 1998; Mori et al., 1999; Poupon et al., 2000]. Comme les modèles locaux représentent l'orientation générale des fibres par voxel, il est intuitif de vouloir connecter entre elles ces orientations. Les Figures 1.10 et 1.11 présentent l'idée générale derrière les algorithmes de tractographie classiques. La tractographie était à l'origine relativement simpliste, utilisant uniquement l'axe principal du tenseur de diffusion par exemple, et s'est graduellement complexifiée en utilisant des modèles locaux plus avancés et en injectant des règles indiquant ce que le tracé peut et ne peut pas faire.

Le résultat d'un algorithme de tractographie est un *tractogramme*, ensemble de milliers de courbes 3D dans le volume cérébral. Les groupes de courbes cohérents entre eux sont parfois nommés *faisceaux*, à ne pas confondre avec les faisceaux anatomiques réellement présents dans le cerveau. Ces courbes 3D, ou *streamlines*, ne représentent pas les axones, mais plutôt un chemin possible entre deux régions potentiellement

### 1.3. RECONSTRUCTIONS PAR TRACTOGRAPHIE

connectées. L’incertitude présente dans cette affirmation vient du fait que cette reconstruction est loin d’être parfaite et contient bien souvent de nombreux faux positifs [Maier-Hein et al., 2017; Schilling et al., 2018b; Thomas et al., 2014]. Ceci découle du processus de décision locale et du fait que les règles ou paramètres n’empêchent pas certains cas limites. De plus, il est difficile de savoir comment la tractographie peut et ne peut pas agir, car beaucoup de détails sont encore inconnus sur l’anatomie du cerveau humain. Les investigations sur l’architecture cérébrale sont complexes, coûteuses et surtout invasives telles que présentées à la Section 1.4. Les mentions d’algorithmes de tractographie présentes dans cette thèse font généralement référence aux algorithmes de tractographie locaux (Section 1.3.1) et plus spécifiquement aux algorithmes de tractographie locaux probabilistes.

#### 1.3.1 Tractographie locale

Les algorithmes de tractographie locale sont les plus intuitifs. Ces approches consistent à simplement initialiser un point de départ dans le cerveau et laisser une courbe se dessiner en suivant les orientations locales à chaque voxel. Lorsqu’on regarde une carte d’orientation, tel qu’illustrée à la Figure 1.10, typiquement l’opération mentale effectuée ressemble à un flot parcourant un champ de vecteurs. Par contre, il est évident qu’une multitude d’approches sont possibles pour effectuer cette tâche et chacune possède son lot de paramètres et d’incertitudes.



### 1.3. RECONSTRUCTIONS PAR TRACTOGRAPHIE

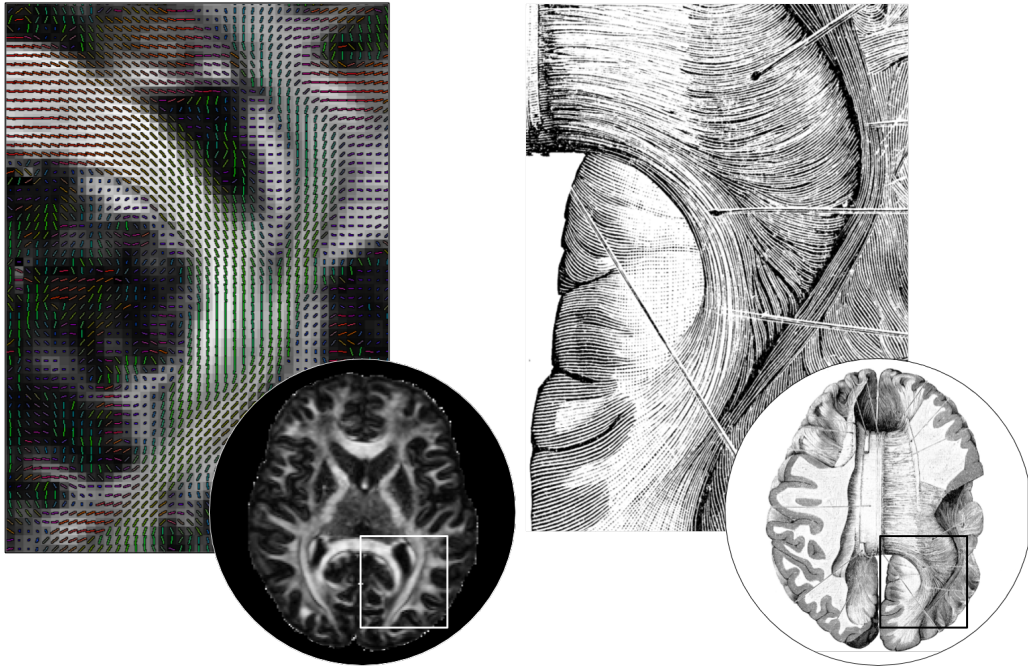


figure 1.10 – À gauche, vue axiale du corps calleux (Splenium). Plutôt que d’illustrer un champ de tenseurs, uniquement la composante principale ( $\lambda_1$ ) du tenseur est affichée. Cette région ne possède qu’une seule population de fibres, ce qui résulte en un champ de vecteurs globalement lisse. Il est visuellement intuitif d’imaginer la structure sous-jacente connectant les deux hémisphères en suivant le flot général de ces vecteurs. L’opération de tractographie a comme tâche de naviguer dans ce champ de vecteurs et de produire des courbes intégrant, sur une certaine distance, cette information d’orientation. À droite, dessin anatomique de [Cunningham \[1906\]](#) illustrant la morphologie du Splenium (agrandi sur la même région) tel qu’observée durant ses travaux de dissections.

Cette liste non exhaustive présente les entrées et paramètres majeurs (et intuitifs) d’un algorithme de tractographie locale.

- **Champs de directions** À chaque voxel une ou plusieurs orientations, ou une fonction de probabilité d’orientation, est nécessaire pour guider les itérations de la tractographie locale. Le concept de champs de directions englobe le résultat de n’importe quel modèle local. À ce jour, l’IRMd est la seule modalité permettant d’obtenir cette information de manière robuste à une résolution acceptable tout en restant accessible dans un contexte clinique. La qualité de



### 1.3. RECONSTRUCTIONS PAR TRACTOGRAPHIE

cette entrée est un facteur majeur pour la tractographie.

- **Masque d’initialisation** Chaque *streamline* possède un point de départ. Le masque d’initialisation peut couvrir tout le cerveau (ou le ruban cortical), ceci permettrait, par exemple, de tenter de reconstruire la connectivité cérébrale dans son entièreté. Par contre, pour adresser une hypothèse plus précise, une région d’intérêt (ROI) peut être définie pour tenter de reconstruire la connectivité d’une seule région en particulier.
- **Masque de tractographie** Les *streamlines* sont contraintes à exister dans un certain espace à l’intérieur du volume cérébral. Ceci est différent du masque d’initialisation. Une fois lancée une *streamline* ne pourra pas traverser un voxel non inclus dans le masque de tractographie. Ce masque inclut, typiquement, la matière blanche et exclut le liquide cébrospinal, simplement parce que le concept de fibre n’existe qu’aux endroits où des axones sont présents et ne fait donc aucun sens dans un tel milieu. Si certains tissus sont affectés par une pathologie le masque de tractographie peut exclure les lésions pour éviter que des *streamlines* traversent un tissu endommagé. La qualité de cette entrée est cruciale pour la tractographie.
- **Courbure maximale** Tel qu’illustré à la Figure 1.12, le processus de tractographie est itératif, pas à pas une *streamline* va avancer dans un champ de vecteurs. L’angle entre deux pas doit être restreint pour éviter les formes aberrantes ou des faisceaux à la morphologie improbable. Ce paramètre aura une influence majeure dans les zones de croisement et permet de contrôler le niveau d’exploration d’un algorithme (particulièrement ceux probabilistes). En effet, l’angle de courbure maximale va empêcher les *streamlines* de soudainement changer de direction, de faire demi-tour ou de *sauter* sur un autre faisceau tel qu’illustré au Chapitre 2 et plus précisément à la Figure 2.6.
- **Longueur maximale et minimale** Il est connu que les axones ne zigzaguent pas aléatoirement ou ne font pas demi-tour au milieu de la matière blanche.

### 1.3. RECONSTRUCTIONS PAR TRACTOGRAPHIE

Il est donc intuitif de limiter la longueur des *streamlines* à quelques centaines de millimètres (200mm-300mm) pour éliminer ce qui est aberrant. De plus, considérant les limitations en termes de résolution d'image, de modèle local et de tractographie, les *streamlines* ayant moins de 10mm-20mm sont difficiles, ou carrément impossibles, à interpréter. Elles ne peuvent donc pas être considérées comme valides et peuvent être automatiquement filtrées. Les axones inter-connectant des gyri adjacents (*fibres en U*) sont plutôt de l'ordre de 40mm-60mm.

- **Pas de propagation** Avant d'avancer pas à pas dans le champ de *directions*, il faut déterminer de quelle taille est le pas. Considérant la résolution d'image et la morphologie de la matière blanche, ce pas est typiquement entre 0.2mm et 1mm. Un pas trop grand et l'algorithme de tractographie pourrait manquer un virage ou un croisement, un pas trop petit et l'algorithme sera grandement ralenti, sera très sensible au bruit et les croisements seront difficiles à traverser. Bien que les résultats de la tractographie peuvent être robustes à plusieurs échelles de pas de propagations, ce paramètre peut être très important pour certaines structures aux configurations complexes et influencera le temps de calcul ainsi que la taille des fichiers.

### 1.3. RECONSTRUCTIONS PAR TRACTOGRAPHIE

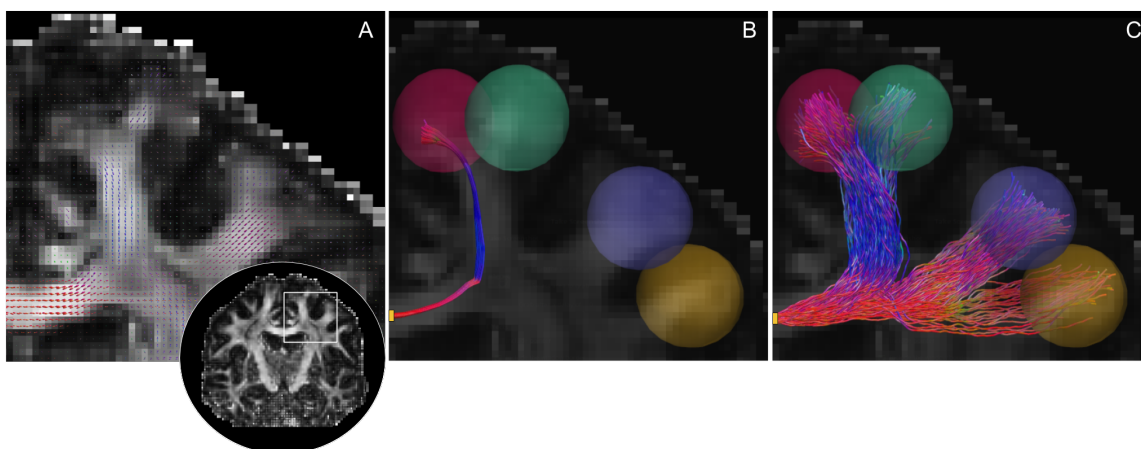


figure 1.11 – Illustration de l’opération de tractographie à partir d’un seul voxel d’initialisation dans le corps calleux (vue coronale). En A, une tranche permettant de voir le champ de fODF que l’opération de tractographie permettra d’explorer. En B, à partir du voxel (boîte jaune) d’initialisation (ligne médiale) 1000 *streamlines* (aléatoirement distribuées) ont été lancées à l’aide d’un algorithme déterministe. Les *streamlines* résultantes, terminant dans l’une des sphères, ont un parcours quasiment identique très conservateur. Étant donné qu’il est connu que le corps calleux peut, et doit, connecter en éventail (fanning) les régions au cortex visible à cette coupe, l’algorithme déterministe procure une reconstruction limitée. En C, l’algorithme probabiliste est initialisé de la même façon et reconstruit partiellement la configuration en éventail attendue. Le chemin le plus dense est celui le plus probable et similaire à B.

#### Approches déterministes

Les algorithmes de tractographie déterministes reposent sur le principe que deux points d’initialisation identiques vont produire deux *streamlines* identiques. La navigation dans le champ de *directions* (peu importe le modèle local) se fait de façon parfaitement prévisible si les entrées sont identiques. La Figure 1.11 illustre ce concept en opposition avec les approches probabilistes. Le choix de direction se fait, typiquement, avec les maxima du modèle local et permet donc de suivre des chemins en ayant, en moyenne, un plus haut niveau de confiance. Ceci permet de réduire la proportion de faux positifs (augmenter la spécificité, et réduire la sensibilité). La reconstruction des principaux faisceaux est assurée, mais leur extension jusqu’au cortex et les configurations difficiles sont parfois ratées.

### 1.3. RECONSTRUCTIONS PAR TRACTOGRAPHIE

#### Approches probabilistes

Les approches probabilistes sont moins intuitives, car elles reposent sur le concept de distribution de probabilité. Contrairement aux algorithmes déterministes, plusieurs possibilités de direction s’offrent à chaque pas. La direction choisie à chaque pas est pondérée grâce au pas précédent ainsi que le modèle local. Les fODF, précédemment présentées à la Section 1.2.2, vont de pair avec les approches probabilistes. En effet, ce modèle représente la fonction de distribution d’orientation de fibres et c’est cette distribution probabiliste du modèle local qui influence la tractographie. La Figure 1.12 illustre la manière dont cette décision est effectuée, tandis que la Figure 1.11 illustre la différence entre un algorithme déterministe et probabiliste à l’échelle d’un faisceau. Le processus de décision permet d’augmenter la capacité de l’algorithme à explorer des chemins plus ou moins probables. Ceci vient au coût d’augmenter la sensibilité, mais de diminuer la spécificité.

#### Ajout de règles basées sur l’anatomie

Les concepts de base de la tractographie peuvent être considérés simples, et même intuitifs, mais de nombreuses extensions ont été développées au fil du temps pour améliorer la qualité des résultats. Par exemple, [Smith et al. \[2012\]](#) et [Girard et al. \[2014\]](#) ont proposé des algorithmes de tractographie intégrant une approche particulière (particles filtering tractography (PFT), Girard) permettant de contourner les obstacles et ainsi permettre d’atteindre de manière plus robuste la matière grise (*anatomically constrained tractography* (ACT), Smith ou *continuous map criteria* (CMC), Girard). L’objectif de ces approches est de se rapprocher de l’anatomie : aucun corps cellulaire n’est présent dans la matière blanche et donc aucun axone ne devrait s’y terminer. Cet *a priori* anatomique est injecté à l’aide de carte probabiliste de tissus et, ainsi, toute *streamline* doit atteindre son objectif pour être considérée valide.

### 1.3. RECONSTRUCTIONS PAR TRACTOGRAPHIE

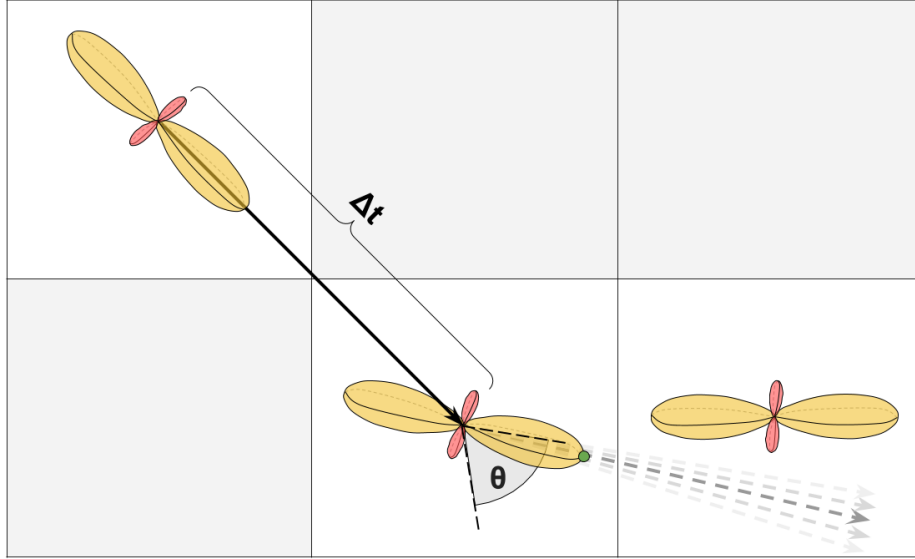


figure 1.12 – Schéma représentant une itération (un pas) du processus de tractographie dans un champ de fODF (3). Une *streamline* initialisée dans le voxel du coin supérieur gauche, effectuant un premier pas dans l'orientation maximale d'une fODF. Le pas d'une longueur  $\Delta t$ , atteint avec un certain angle le voxel central inférieur. Pour éviter des changements brusques de direction (limiter la courbure de la *streamline*), le pas suivant ne pourra pas être en dehors d'un cône ayant un angle d'ouverture  $\theta$ . À l'intérieur de ce cône, le choix de la direction suivante s'effectuera à l'aide de l'amplitude de la fODF à ce voxel. Un algorithme déterministe choisira systématiquement le maximum local présent dans le cône, tandis qu'un algorithme probabiliste choisira une direction à l'aide d'une probabilité pondérée par l'amplitude de la fODF. Les flèches d'intensité variable montrent les directions qui seront possiblement choisies au prochain pas par un algorithme probabiliste. Celles claires étant moins probables que celles foncées, un algorithme déterministe choisira un maxima local (flèche foncée au centre).

Un autre exemple d'extension a été proposé par [St-Onge et al. \[2018\]](#) pour tenter de corriger le biais gyral. Le biais gyral est la tendance de la tractographie à se terminer à la couronne des gyri plutôt que de se disperser uniformément sur toute la surface intérieure des gyri (voir Chapitre 2, plus précisément Section 2.2.2). Le but de cette contribution est d'augmenter la qualité des tractogrammes à la frontière entre la matière blanche et la surface corticale. L'intégration de surface ayant une plus haute résolution qu'un volume de diffusion permet de mieux modéliser la trajectoire

### 1.3. RECONSTRUCTIONS PAR TRACTOGRAPHIE

des *streamlines* près des circonvolutions corticales. Encore une fois, l’objectif est de se rapprocher de ce qui est connu de l’anatomie et de rejeter les trajectoires ne respectant pas certains critères liés à la dispersion à la frontière entre la matière blanche et la surface corticale.

Ces deux exemples ayant émergé de notre laboratoire ne sont évidemment pas une liste exhaustive, mais représentent, en essence, le type d’*a priori* ajouté aux algorithmes de tractographie. Il s’agit typiquement de règles devant être respectées à l’échelle du cerveau et permettant de contraindre les aspects non désirables et/ou d’encourager ceux qui sont connus. Par contre, l’ajout de ces règles se fait bien souvent à l’aide d’information obtenue d’autres modalités nécessitant leurs propres algorithmes complexes. Ceci ajoute parfois de nouveaux types d’erreurs jusqu’à présent inexistants ou inconnus (voir Chapitre 2).

#### 1.3.2 Tractographie globale

La seconde grande catégorie d’algorithme de tractographie ne repose pas sur le concept d’itération où des pas sont pris un à un dans un champ de directions. La tractographie globale repose sur le principe que l’entière d’une *streamline* est évaluée à l’aide d’une fonction de coût qui doit être minimisée. Évidemment de nombreuses approches ont été proposées [Christiaens et al., 2015; Fillard et al., 2009; Kreher et al., 2008; Mangin et al., 2013; Neher et al., 2012; Reisert et Kiselev, 2011; Teillac et al., 2017], mais l’objectif est de générer un ensemble de *streamlines*, un tractogramme, optimal selon les paramètres d’entrées. La fonction de coût prend non seulement en compte le signal de diffusion, mais peut prendre en compte des contraintes sur la morphologie ou encourager les connexions *directes*, les plus courtes, dans un espace géodésique.

Ces approches, souvent plus récentes, sont généralement considérées comme supérieures aux approches locales. En effet, plusieurs ont cru que ces méthodes surmonteraient certaines limitations dues à leur complexité. Hélas, de nombreux biais (voir Chapitre 2) semblent inhérents à tous les algorithmes de tractographie et les avantages proclamés des approches globales sont parfois discutables. La complexité des

## 1.4. MÉTHODES DE VALIDATIONS

équations mathématiques, le grand nombre de paramètres d'entrée non intuitifs et l'énorme temps de calcul font que ce type d'algorithme est encore rare et plus ou moins applicable en milieu clinique.

## 1.4 Méthodes de validations

Du point de vue de l'imagerie médicale, les méthodes de validations telles que la coloration histologique ou les dissections post-mortem sont considérées comme des «*vérités terrain*» à nos reconstructions par tractographie. Par contre, il est important de connaître ces méthodes et leurs propres limitations pour réellement comprendre la complexité de la tâche qu'est la cartographie du cerveau humain. Toutes les méthodes de validation anatomique présentées ci-dessous sont effectuées post-mortem, car elles détruisent les tissus étudiés. Ces méthodes de validations utilisées depuis des années ont permis d'obtenir des confirmations de l'existence ou de l'absence de certains faisceaux. Un seul récit n'est pas suffisant, mais la somme des études de validation a permis au cours des dernières décennies (ou plutôt du dernier siècle) d'avoir un ensemble de connaissances anatomiques fiables entourant les principaux faisceaux de la matière blanche ou de la connectivité des noyaux gris centraux.

### 1.4.1 Dissection post-mortem

La manière la plus ancienne d'investiguer l'anatomie cérébrale est la dissection manuelle post-mortem ou *blunt dissection* [Dejerine et Dejerine-Klumpke, 1895; Déjerine et Dejerine-Klumpke, 1901]. En raclant/grattant les tissus pour atteindre la matière blanche profonde et en suivant l'orientation générale de la texture fibreuse, la position et la géométrie des faisceaux peuvent être exposées [Ludwig et Klingler, 1956]. Par contre, les dissections exploratoires sont difficiles, car la procédure détruit les faisceaux dans le plan perpendiculaire au raclage/grattage. L'orientation et position approximative doivent être planifiées avant d'entreprendre la dissection pour exposer correctement le faisceau d'intérêt [Benedictis et al., 2016; Hau et al., 2017; Maffei et al., 2018] (voir Figure 1.13). Il est ensuite difficile de lier les observations faites durant la dissection aux données d'imagerie médicale [De Benedictis et al., 2018; Sa-

#### 1.4. MÉTHODES DE VALIDATIONS

[rubbo et al., 2019](#)], ce qui explique pourquoi cette approche est parfois perçue comme qualitative. Les tendances macroscopiques peuvent atteindre un consensus, mais les variations dans les préparatifs de la procédure (et la dissection elle-même) rendent difficiles d'obtenir des définitions anatomiques exactes.

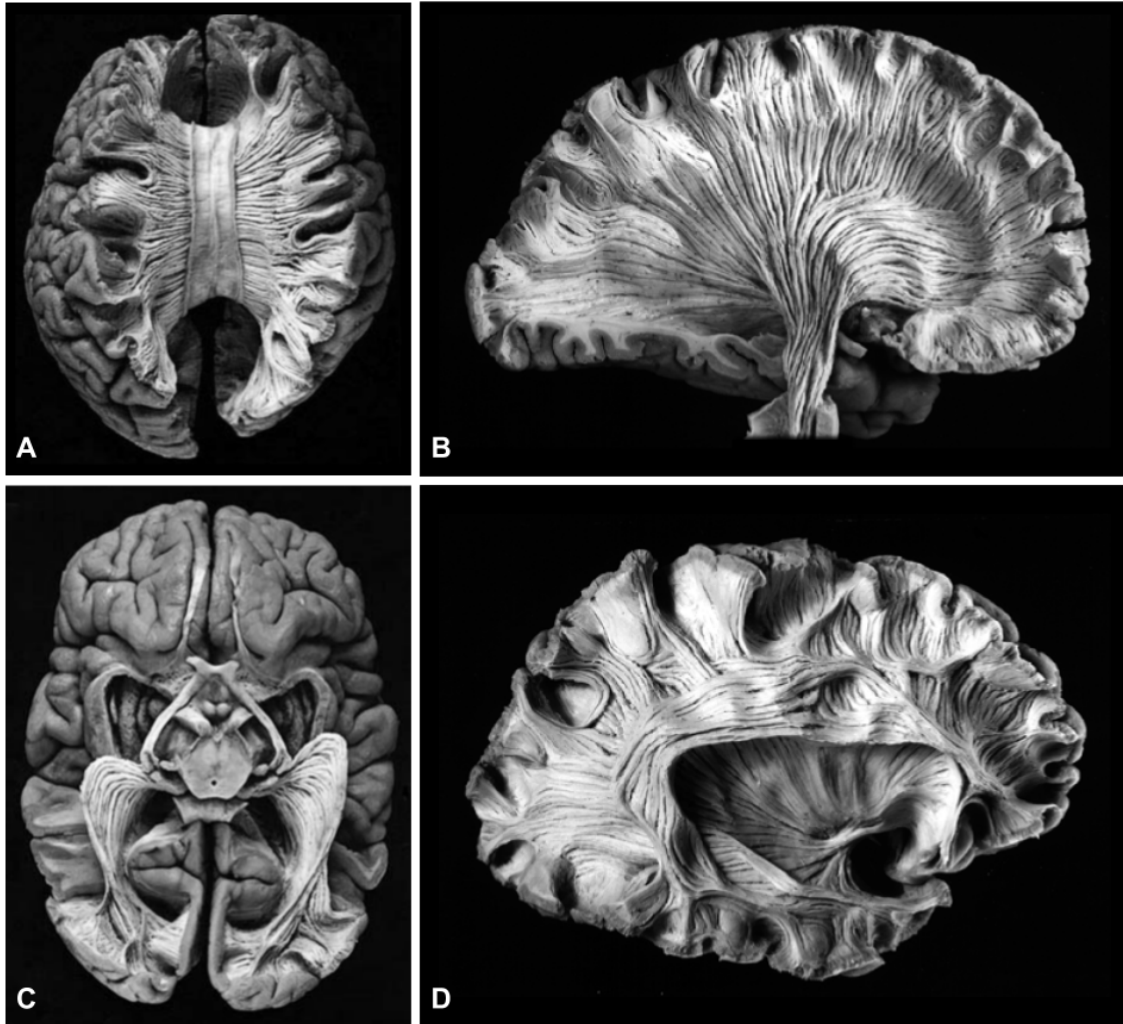


figure 1.13 – Dissection de différentes structures d'intérêt de la matière blanche. Le corps calleux (vue supérieure, A), la couronne rayonnante ou *corona radiata* (vue latérale droite, B), radiations optiques (vue inférieure, C) et le faisceau arqué (vue latérale de droite, D). Le processus de préservation du cerveau affecte la texture fibreuse de la matière blanche et permet d'isoler plus facilement la structure plutôt cohérente des faisceaux. Figure adaptée de [\[Williams et al., 1999\]](#).



## 1.4. MÉTHODES DE VALIDATIONS

### 1.4.2 Coloration histologique

La coloration histologique est une opération complexe où les tissus cérébraux sont préservés en utilisant des agents de fixation, coupés en minces tranches et «colorés» pour faciliter les observations au microscope. L’origine de ce type d’approche remonte à plus d’une centaine d’années et une variété impressionnante de techniques ont été proposées depuis [Alturkistani et al., 2016; Swanson, 2000; Záborszky et al., 2006]. Ces techniques de coloration sont utilisées pour augmenter le contraste entre les différents types de cellules, tel qu’illustré dans la Figure 1.14. Cependant, selon les types d’agents de fixation, les méthodes de découpe, le processus de coloration et les algorithmes utilisés pour l’analyse, les structures d’une même région peuvent être décrites de façon différente. La manière d’effectuer toutes ces étapes fait encore l’objet de débat, ou du moins d’interprétations variables, même pour les laboratoires de pointe. Étant donné que l’analyse se fait sur une seule tranche à la fois, la reconstruction sur plusieurs tranches est un défi algorithmique de taille. La coloration histologique est un outil précieux pour la validation de modèles de microstructures.

#### 1.4. MÉTHODES DE VALIDATIONS

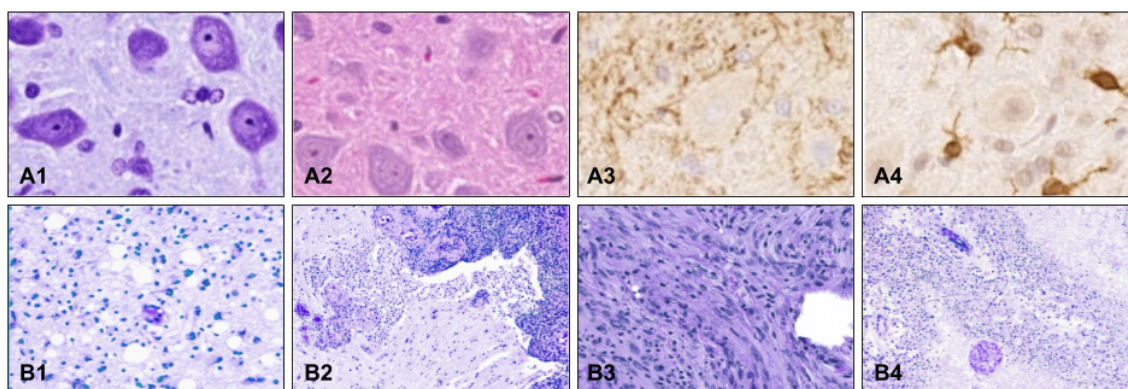


figure 1.14 – Exemples de différents types de coloration histologique et de différents types de tissus. La rangée A montre 4 marqueurs chimiques permettant d’observer différents types de cellules, tandis que la rangée B montre 4 types de tissus avec le même colorant (coloration hématoxyline et à l’éosine). Ces colorations permettent de distinguer des tissus anormaux ou de mesurer la densité d’un type de cellules en particulier. Les types de coloration montrés en A sont (gauche à droite) : H&E, Nissl stain, GFAP et Iba1. Les types de cellules montrés en B sont issus d’histologies de tissus cancéreux (gauche à droite) : astrocytome diffus, méningiome invasif, glioblastome et méningiome fibroblastique. Figure adaptée de [Lohrke et al., 2017] et [Nilsson et al., 2018]

Plus spécifiquement à l’analyse de la connectivité cérébrale, le traçage axonal est accompli en injectant un agent colorant conçu pour se déplacer le long des axones [Kristensson et Olsson, 1973; Lavail et Lavail, 1972; Saleeba et al., 2019]. L’injection se fait, typiquement, dans une région d’intérêt du cortex, le corps cellulaire des neurones absorbe l’agent et celui-ci va être transporté le long de l’axone révélant ainsi son parcours [Schmahmann et Pandya, 2006; Schmahmann et al., 2007; Yeterian et al., 2012], voir Figure 1.15. Le traceur peut agir de manière antérograde (du corps cellulaire aux terminaisons synaptiques) ou rétrograde (des terminaisons synaptiques au corps cellulaire). Localement, des algorithmes d’analyse de textures permettent d’obtenir des informations d’orientation, ce qui permet de valider les modèles locaux obtenus par l’imagerie de diffusion. Ces agents colorants nécessitent toute une période d’attente pour que les molécules de l’agent soient transportées. Ensuite, le cerveau peut être tranché et analysé. Encore une fois, ces étapes peuvent varier selon le laboratoire. La complexité de ces étapes et le grand nombre de variations dans les protocoles de vali-

#### 1.4. MÉTHODES DE VALIDATIONS

dation peuvent grandement influencer la connectivité observée. Chaque protocole de validation ne montre probablement qu'une représentation partielle de la connectivité réelle [Saleeba et al., 2019].

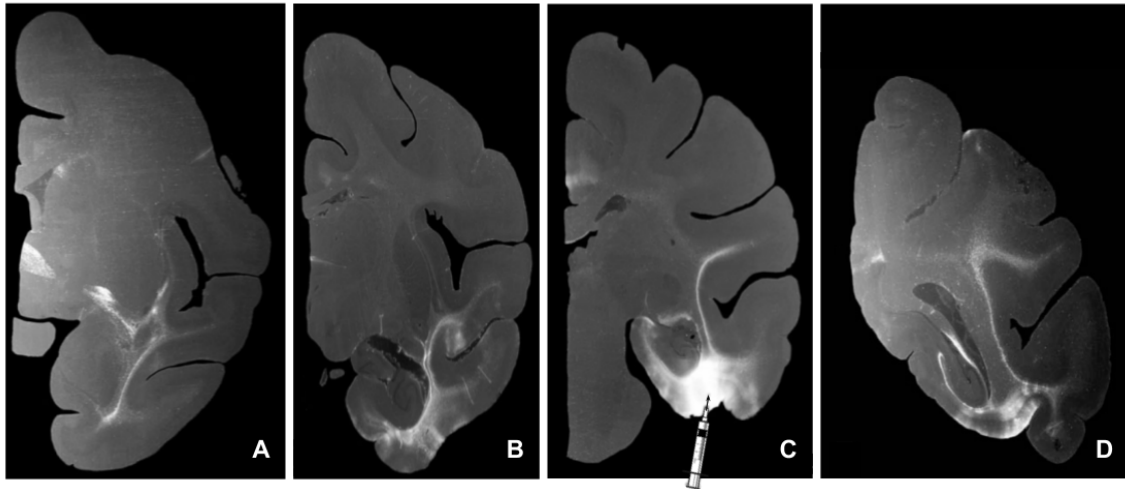


figure 1.15 – Multiples tranches montrant la propagation d'un traceur axonal jointe à une imagerie par autoradiographie [Meyer et Meyer, 1973] provenant d'un cerveau de singe rhésus. Le site d'injection est visible en C (lobe temporal), la connectivité de cette région est exposée grâce au procédé d'autoradiographie. Figure adaptée du livre *Fiber pathways of the brain* [Schmahmann et Pandya, 2006].

## 1.4. MÉTHODES DE VALIDATIONS

### 1.4.3 Approches modernes

#### Imagerie par lumière polarisée

L'imagerie par lumière polarisée nécessite une découpe du cerveau en tranches minces, cette méthode permet la reconstruction d'un champ de vecteurs dense plutôt qu'une information de connectivité. En faisant varier l'angle de polarisation d'une source de lumière, la structure fibreuse de la tranche peut être analysée. Chaque pixel contient l'information d'orientation [Axe et al., 2000] (voir Figure 1.16A-B). Cette information ne permet pas de confirmer la présence ou l'absence d'une connexion, mais permet plutôt de valider les modèles locaux obtenus par l'imagerie de diffusion [Axe et al., 2011a, 2016; Mollink et al., 2017; Reckfort et al., 2015].

#### Tomographie en cohérence optique

La tomographie en cohérence optique permet d'analyser la structure fibreuse d'une tranche à l'aide d'une source de lumière pouvant varier sa longueur d'onde. Cette méthode permet d'imager à haute résolution une tranche mince et, suite à une analyse de texture, obtenir de l'information sur l'orientation locale des fibres (voir Figure 1.16B). Cette information peut ensuite servir à valider les modèles locaux obtenus de l'imagerie de diffusion [Wang et al., 2015, 2011]. Par contre, l'analyse est limitée par l'opération de découpe et l'analyse des longues connexions de la matière blanche profonde est toujours difficile.

## 1.4. MÉTHODES DE VALIDATIONS

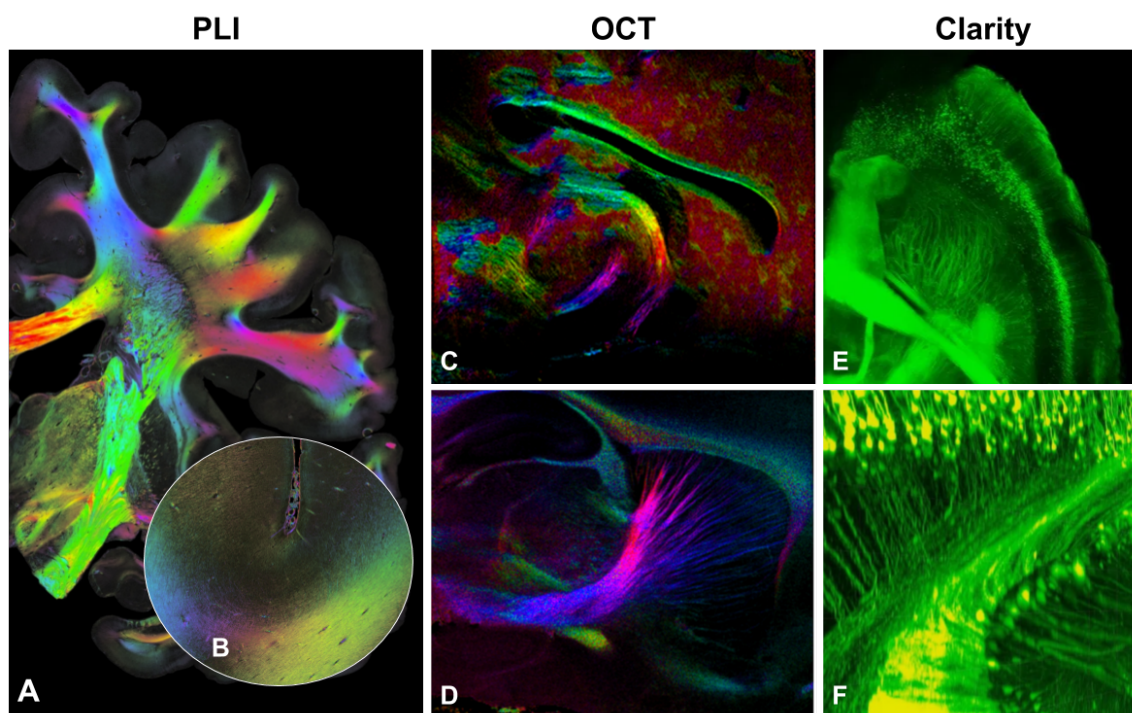


figure 1.16 – Présentation de 3 méthodes de validation modernes basées sur la microscopie. En A, une tranche coronale d'un cerveau humain acquise à l'aide de lumière polarisée. En B, une région agrandie d'un gyrus possédant des fibres en U ayant une grande courbure. La coloration représente l'orientation locale des fibres, ceci est une image composite reconstruite à l'ordinateur. C et D montrent une région entourant les noyaux gris centraux d'un cerveau humain. La méthode de tomographie en cohérence optique permet d'obtenir l'orientation locale des fibres qui est représentée par une coloration à l'ordinateur. En E et F, un morceau 3D de cerveau de souris transgénique où les tissus gras sont transparents et où les neurones ont été colorés puis imagés à l'aide d'un microscope. Le résultat est une image en 3D. Figure adaptée de [Axer et al., 2011b; Chung et Deisseroth, 2013; Wang et al., 2015]

### Clarity

Clarity peut être décrit comme une méthode de traçage axonal en 3D. L'approche nécessite un bloc de tissus plutôt qu'une tranche mince, ce qui facilite l'analyse sur des échantillons de plus grande taille [Chang et al., 2017; Chung et Deisseroth, 2013; Morawski et al., 2018; Spence et al., 2014]. Une procédure chimique rend les tissus gras transparents et expose les types de cellules désirés pour l'analyse (e.g axones)

## 1.5. ANALYSE DE TRACTOGRAMME

(voir Figure 1.16C). Le résultat étant un volume montrant le parcours d'axones, l'étape d'analyse peut aussi bien tenter de reconstruire la connectivité ou l'orientation générale présente dans un bloc. Par contre, les tailles de l'échantillon sont à l'échelle d'à peine un centimètre, encore bien loin d'être à l'échelle des faisceaux de la matière blanche.

Les trois méthodes précédemment mentionnées contiennent leurs lots de limitations méthodologiques (fixations, découpes, variabilité expérimentale, etc.) et défis algorithmiques (analyse de texture, recalage, etc.). Ceci rend difficile d'obtenir des observations cohérentes sur la neuroanatomie et sur l'atteinte d'un consensus sur la neuroanatomie qu'elles tentent d'éclaircir. C'est à l'aide de plusieurs publications au fil du temps qu'une tendance peut être observée et seulement après que des *a priori* pour la tractographie peuvent être formulés.

## 1.5 Analyse de tractogramme

Une fois la reconstruction d'un tractogramme terminée, peu importe l'approche utilisée, les méthodes d'analyse restent typiquement les mêmes. Lorsque des hypothèses concernent un processus cognitif lié à une connexion précise, l'approche de segmentation, ou dissection, de faisceaux est utilisée. Cette approche permet d'isoler un ou des groupes de *streamlines*, ayant une cohérence fonctionnelle et/ou une morphologie connue grâce à des méthodes de validation (voir Section 1.4).

L'approche par connectome, ou *connectomics* est rarement utilisée pour étudier un processus cognitif précis, mais plutôt pour quantifier ou corrélérer l'ensemble de la connectivité structurelle. L'approche par connectome nécessite un atlas de régions corticales, une parcellisation, et un tractogramme couvrant l'entièreté du cerveau. L'objectif est d'évaluer les connexions entre toutes les paires de régions corticales et de les rapporter sous la forme d'une matrice telle qu'illustrée à la Figure 1.18.

## 1.5. ANALYSE DE TRACTOGRAMME

### 1.5.1 Dissection virtuelle

Les connaissances acquises grâce aux approches de validation ont permis de définir bon nombre de faisceaux anatomiques présents chez l’humain. Un peu comme l’architecture corticale où les noyaux centraux ont été étudiés et sous-divisés selon leurs rôles et fonctions, les connexions de la matière blanche peuvent être classifiées de manière similaire. Bien que les définitions proposées par les experts en neuroanatomie ne font pas toujours consensus (voir Section 2.3.2), il est possible de les interpréter et de créer des règles cohérentes pour isoler certains faisceaux.

Ce type d’approche permet de mettre l’accent sur les faisceaux dont la contribution fonctionnelle est connue et où la variabilité anatomique à l’échelle de la population est relativement faible. Il existe plusieurs façons de quantifier les faisceaux extraits à l’aide d’une dissection virtuelle : rapporter le volume, le nombre de *streamlines* ou même la valeur moyenne d’une métrique à l’intérieur du faisceau. Cette quantification spécifique à l’approche par faisceau se nomme *tractométrie* ou *profilométrie* [Cousineau et al., 2017; Yeatman et al., 2012, 2018]. Par exemple, en segmentant un faisceau en particulier, au cours d’une étude longitudinale, il est possible d’étudier son évolution en traçant son volume en fonction du temps. Certains biais et défis liés aux dissections virtuelles sont présentés dans le Chapitre 2.

### Approches manuelles

Une dissection, ou segmentation, manuelle se fait en utilisant des règles définissant les points de passage, la position attendue des terminaisons corticales et en excluant les régions où un faisceau en particulier ne peut exister, un faisceau peut être isolé d’un tractogramme [Bayrak et al., 2019; Catani et De Schotten, 2008; Catani et al., 2002; Chenot et al., 2019; Hau et al., 2016]. Cette opération peut être effectuée par un expert en neuroanatomie suivant un plan clairement défini à l’avance ou par inspections visuelles pour isoler un groupe de *streamlines* ayant une morphologie similaire à ce qui est attendu. Par exemple, dans le cas d’une recherche clinique ayant une grande cohorte il est nécessaire de suivre un plan clairement défini [Catani et al., 2007; Groeschel et al., 2014; Lee Masson et al., 2017; Masson et al., 2018], tandis que pour une situation en médecine personnalisée dans le cas d’une chirurgie l’approche visuelle



## 1.5. ANALYSE DE TRACTOGRAMME

pourrait suffire [Caverzasi et al., 2016; Essayed et al., 2017; Fernández-Miranda et al., 2008; Pujol et al., 2015].

L'exécution d'une tâche de dissection virtuelle repose sur des repères anatomiques parfois difficiles à identifier [Rheault et al., 2020]. Même dans le cas où un plan de dissection a été clairement établi à l'avance, il est difficile de garantir que la tâche sera effectuée correctement par un individu. Ce problème de variabilité est discuté plus en détail au Chapitre 3.

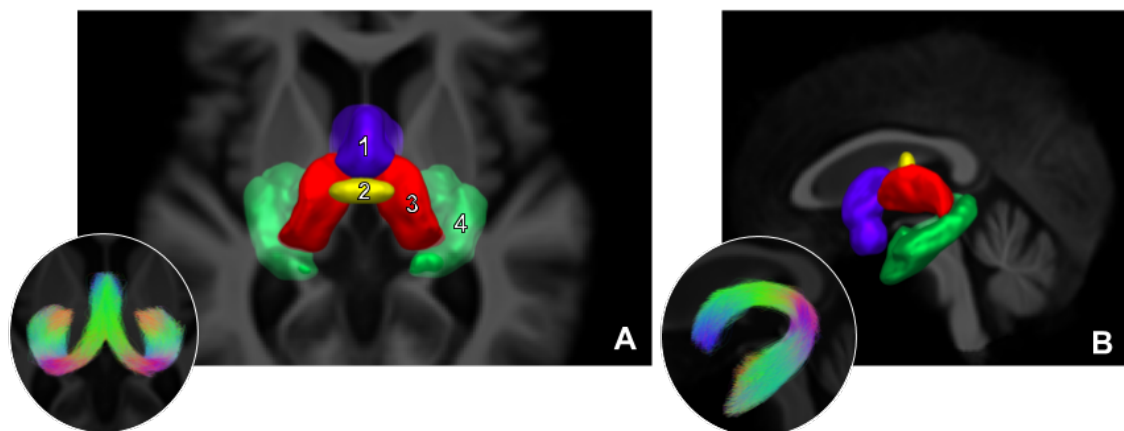


figure 1.17 – Vue axiale (A) et vue sagittale (B) des régions d'intérêt utilisées pour la segmentation manuelle du fornix, faisceau commissural inter-connectant les hippocampes liés à différents processus de la mémoire. Corps mamillaires en mauve (inclusion, 1), corps du fornix en jaune (inclusion, 2), deux thalami en rouge (exclusion, 3) et deux hippocampes en vert (inclusion, 4). En vignette, la morphologie générale du fornix une fois segmenté.

### Approches automatiques

Les approches automatiques ont l'objectif d'extraire le même type de faisceau, mais en définissant les règles de manière à permettre l'exécution d'un algorithme sans intervention humaine. L'approche automatique la plus intuitive est l'utilisation d'atlas divisant le volume cérébral, typiquement en se basant sur des repères anatomiques, et utiliser celui-ci afin de définir le plan de dissection en termes de commande exécutée par un logiciel [Wassermann et al., 2016]. Cette approche s'apparente à celle décrite ci-haut et nécessite une connaissance de l'anatomie tout autant importante. Définir un



## 1.5. ANALYSE DE TRACTOGRAMME

faisceau de manière robuste n'est pas une opération triviale [Catani et De Schotten, 2008; Mandonnet et al., 2018; Panesar et Fernandez-Miranda, 2019].

Certains algorithmes reposent sur des algorithmes de regroupement basés sur la similarité des terminaisons ou la morphologie des *streamlines* pour ensuite associer ces groupes à des faisceaux anatomiques connus [Garyfallidis et al., 2018; Guevara et al., 2012; O'Donnell et al., 2017; Yoo et al., 2015]. Le chapitre 4 propose une extension d'un algorithme de segmentation automatique reposant sur la similarité à plusieurs modèles de faisceaux (Recobundles proposé par Garyfallidis et al. [2018]).

### 1.5.2 Connectome

L'approche par connectome [Hagmann, 2005] est toute autre, l'absence de règles dans la matière blanche facilite en quelque sorte la tâche, mais transfère la difficulté sur le choix de parcellisation corticale [Sotiropoulos et Zalesky, 2019]. De plus, la construction d'une matrice de connectome est typiquement effectuée avec moins de connaissance *a priori*, ce qui la rend particulièrement sujette aux biais. Ces biais, décrits en plus de détails au Chapitre 2, rendent l'interprétation difficile et distancient en quelque sorte l'approche par connectome des processus cognitifs. Par contre, en résumant la connectivité cérébrale globale à une matrice pouvant prendre la forme d'un réseau mathématique, les possibilités d'analyses sont plus variées.

Plusieurs études impliquant des matrices de connectome ont été effectuées pour classer des individus atteints de pathologie cérébrale. L'objectif à long terme est généralement de faciliter le diagnostic dans le futur [Alexander-Bloch et al., 2013; Fernandes et al., 2015; Griffa et al., 2013; Ingahalikar et al., 2014]. Certains groupes ont aussi tenté d'investiguer les troubles du développement ayant une influence plus généralisée sur le développement tel que le trouble déficitaire de l'attention ou sur les différents degrés du spectre de l'autisme [Griffiths et al., 2016; Valk et al., 2015].

## 1.6. CONCLUSION

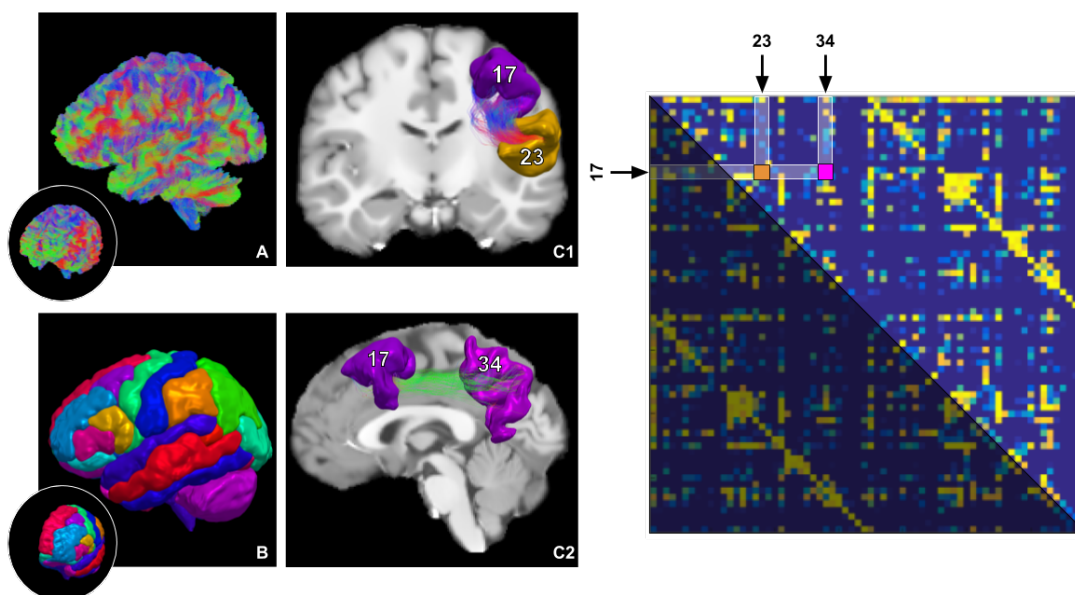


figure 1.18 – Composantes nécessaires à la construction d’une matrice de connectome, soit un tractogramme (A) et une parcellisation corticale (B). Chaque région est utilisée pour la sélection de *streamlines*, en C et D les régions #17, #23 et #34 permettent d’isoler des *streamlines*. Le nombre de *streamlines* est ensuite rapporté dans la matrice (symétrique) aux coordonnées (17, 23) et (17, 34). Une parcellisation corticale avec, par exemple, 100 étiquettes produira une matrice de connectome de dimension 100x100.

## 1.6 Conclusion

Les efforts liés à l’étude du *Connectome* humain sont de l’ordre de ceux du séquençage du *Génome* humain. Des centaines de millions de dollars sont investis annuellement par divers gouvernements un peu partout dans le monde pour mieux comprendre le fonctionnement et l’architecture du cerveau grâce aux techniques de pointe de l’IRM.

L’ampleur de la tâche force la collaboration entre plusieurs domaines scientifiques. L’existence de groupes de recherches multidisciplinaires est donc une nécessité. Le scanner par résonance magnétique et les séquences d’acquisition d’IRMd sont d’une complexité impressionnante, et cela même pour un chercheur ayant une spécialisation

## 1.6. CONCLUSION

en physique. Les différents modèles locaux nécessitent non seulement une connaissance des mathématiques, mais aussi du processus physique de diffusion et parfois même de la biologie ou de l'anatomie sous-jacente. La tractographie en tant que telle est un défi autant mathématique qu'algorithmique et la taille des bases de données modernes ne fait que complexifier les étapes de traitement de données. L'analyse et l'interprétation des résultats nécessitent bien souvent des connaissances précises en statistiques, en anatomie et en neurosciences.

L'objectif à long terme de la tractographie est de suffisamment améliorer la qualité des reconstructions virtuelles pour que les tractogrammes résultants de la tractographie soient dépourvus de biais majeurs. Pour atteindre un tel objectif, des efforts multidisciplinaires intensifs devront être mis en place et ultimement unifier les travaux de plusieurs groupes. Bien que les reconstructions virtuelles par tractographie ne sont qu'une approximation du véritable *Connectome* [Sporns et al., 2005], il s'agit d'un outil précieux pour les neurosciences modernes. L'intérêt grandissant pour l'IRMd et la tractographie est en soi un défi pour ce récent domaine. Le développement de méthodes plus simples à utiliser, ayant une validité accrue, des résultats plus robustes et produisant des conclusions cohérentes pouvant être répliquées et propulsera la tractographie vers une utilisation plus courante en recherche et en clinique.

“I don’t believe in astrology ; I’m a  
Sagittarius and we’re skeptical.”

---

— Arthur C. Clarke, *Astronomy Now*.

## Chapitre 2

# Perceptions erronées, biais cachés et défis modernes de la tractographie

### Résumé

Le cerveau humain est un réseau complexe et organisé, où chaque connexion entre régions est le résultat non pas d'axones s'interceptant les uns avec les autres, mais bien à l'aide de millions d'axones densément entassés et bien structurés. Les reconstructions provenant de la tractographie par imagerie par résonance magnétique de diffusion ne sont qu'une tentative de capturer la complexité de ce réseau à l'échelle macroscopique. Cette publication procure une vue d'ensemble des perceptions erronées, des biais et pièges présents dans les reconstructions de faisceaux et de connectomes basées sur la tractographie. L'objectif de cette revue n'est pas de décourager les lecteurs, mais bien de les informer sur les limitations méthodologiques des outils utilisés par les chercheurs du domaine. Ceci permettra donc de mieux se concentrer sur les forces de la tractographie et d'encourager une interprétation adéquate des résultats. De plus, cette revue procure une liste de problèmes ouverts et de défis qui pourraient être résolus par la prochaine génération d'étudiants et leurs projets de recherche.

## Contributions de la publication

- Vulgarisation des limitations liées au traitement de données qui affectent spécifiquement la tractographie.
- Agrégations de nombreuses limitations algorithmiques observées dans le laboratoire (mais jamais formellement publiées).
- Illustration de la variabilité des résultats selon les choix algorithmiques, d’outils, atlas ou de paramètres.

## Contributions des auteurs

- Écriture du manuscrit et création des figures (Francois Rheault)
- Création de la figure 2.7 liée au biais gyral (Etienne St-Onge)
- Création des données simulées et de l’image 2.4 liée au “*wall effect*” (Alex Valcourt-Carron)
- Correction et relecture du manuscrit (Philippe Poulin, Maxime Descoteaux)

## Commentaires

L’origine de cette publication était en fait une commande pour un chapitre de livre (*Handbook of tractography*). Ce texte didactique avait comme objectif de «révéler» les limitations, biais et pièges de la tractographie aux chercheurs débutants dans ce domaine. Une discussion informelle sur ce chapitre a eu lieu à ISMRM Paris en 2019 et un plan d’écriture a été décrit à l’aide de tous les membres du laboratoire. Un premier brouillon a été écrit, avec papiers et crayons, dans les forêts du Yosemite et Sequoia National Parks. En août 2019, une demande de publication sur le même sujet pour le journal *Neural Engineering* a été envoyée à Maxime Descoteaux. Ce brouillon a rapidement été converti en publication pour ce journal. Ce chapitre est la soumission à l’édition spéciale *Non-invasive brain imaging* de ce journal (soumise 1er novembre 2019, en révision)

## Common misconceptions, hidden biases and modern challenges of dMRI tractography

Francois Rheault<sup>1</sup>, Philippe Poulin<sup>1</sup>, Alex Valcourt Caron<sup>1</sup>, Etienne St-Onge<sup>1</sup>,  
Maxime Descoteaux<sup>2</sup>,

<sup>1</sup>Sherbrooke Connectivity Imaging Laboratory (SCIL), Université de Sherbrooke, Canada

<sup>2</sup>Groupe d'Imagerie Neurofonctionnelle, IMN, UMR5293, Université de Bordeaux, France

**Keywords:** Diffusion MRI, White Matter, Tractography, Tractometry, Connectivity

### Abstract

The human brain is a complex and organized network, where the connection between regions is not achieved with single neurons crisscrossing each other but rather millions of densely packed and well-ordered neurons. Reconstruction from diffusion MRI tractography is only an attempt to capture the full complexity of this network, at the macroscale. This review provides an overview of the misconceptions, biases and pitfalls present in structural white matter bundle and connectome reconstruction using tractography. The goal is not to discourage readers, but rather to inform them of the limitations present in the methods used by researchers in the field in order to focus on what they can do and promote proper interpretations of their results. It also provides a list of open problems that could be solved in future research projects for the next generation of PhD students.

## 2.1 Introduction

Diffusion magnetic resonance imaging (dMRI) tractography is the only non invasive tool to probe the white matter (WM) architecture of the brain. The wide range of processing related to dMRI tractography, the variety of "tools/software" and the large number of laboratories with different backgrounds create a wealth of options difficult to untangle for new researchers.

## 2.1. INTRODUCTION

The nature of the field makes multidisciplinary a necessity, from physics and mathematics to computer science and biomedical engineering and then to neuroanatomy and medicine (and much more), all are essential. The level of specialization required at each and every step of a project means that numerous misconceptions, biases and pitfalls can arise along the way [Calamante, 2019; Jbabdi et Johansen-Berg, 2011; Jones et al., 2013; Maier-Hein et al., 2017; O'Donnell et Pasternak, 2015; Sotiropoulos et Zalesky, 2019].

However, the purpose of this review is only to present and acknowledge known misconceptions/biases/pitfalls related to dMRI tractography that lead to inconclusive results and makes it challenging to achieve anatomically-relevant interpretations. Being aware of theses challenges is crucial for new students and researchers if we want them solved in the near future.

The popularity of dMRI tractography in clinical research has shown a relentless growth since its humble beginnings, from being an obscure methodological approach to being in the spotlight at clinical or neurosurgical conferences and MRI vendor infographics and publicity. The growing popularity of dMRI tractography (see Figure 2.1) and the slew of tools and software available makes it more accessible than ever. However, as the field is switching its focus from methodological publications towards more clinical applications, specific misconceptions/biases/pitfalls often well-known by those who developed tools and software, are lost to the general users.



## 2.1. INTRODUCTION

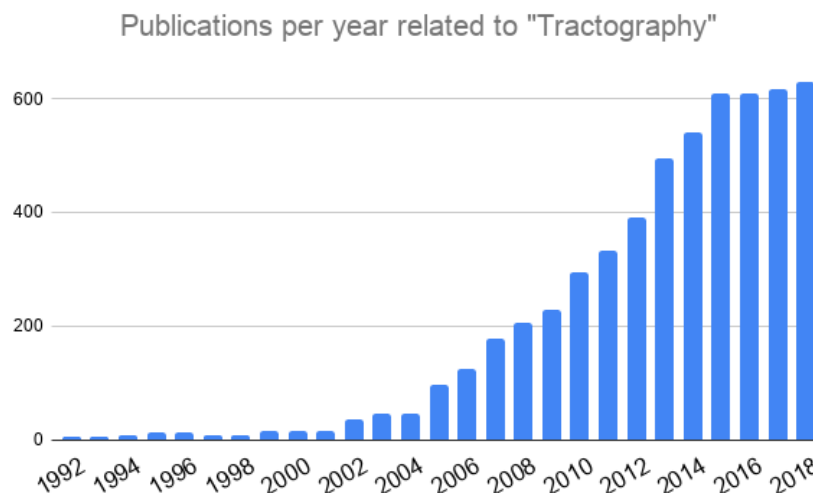


figure 2.1 – Occurrence of the keyword *Tractography* on the digital repository PubMed Central. It is now more popular than ever. In the early days of tractography, the keyword was not even present, until it took off in the 2000s. These numbers are not even taking into account the publications calling it DTI.

This work will provide an overview of the challenges faced by the field. These challenges are not always biases or pitfalls, but they are all source of misunderstanding and lead to misinterpretations. First, we review the various problems that arise due to local decisions (local models, masking, tractography algorithms, etc.) and their impact on the production of individual streamlines (or tracks). Then, we tackle WM bundle reconstruction and segmentation and finally the connectome reconstruction as a whole. Finally, we review common problems related to bundle segmentation and connectomics while taking previously mentioned biases and pitfalls into account.

The reader is expected to have a good grasp of local modelling, tractography algorithms, diffusion-related measures, manual and automatic WM bundle segmentation as well as being familiar with typical clinical study design involving dMRI tractography. [Descoteaux, 2015; Johansen-Berg et Behrens, 2013]

## 2.2 Tractography versus underlying anatomy, what is and what isn't ?

One of the main challenges of dMRI tractography resides in the fact that it is expected to reconstruct the connectivity at the entire brain level using only local diffusion information. Even global tractography methods ultimately rely on local information patched together [Daducci et al., 2016; Jeurissen et al., 2017; Maier-Hein et al., 2017]. This conundrum cannot be easily solved and often leads to streamlines that look valid but have no basis in the underlying anatomy, or streamlines that obviously look invalid, but until verified cannot be discarded. While dMRI tractography can indeed reconstruct correctly known WM anatomy [Benedictis et al., 2016; Chenot et al., 2019; Hau et al., 2017; Maffei et al., 2018], this problem makes it difficult to use tractography for exploratory investigations. This section will detail some of the known biases and pitfalls when it comes to tractogram reconstruction. Note that despite numerous examples, this list is not exhaustive and does not cover all existing algorithms.

### 2.2.1 Local signal limitations

Despite all the improvements made to dMRI tractography [Jeurissen et al., 2017] and local modelling [Dell'Acqua et Tournier, 2019] over the last few years, it is important to reiterate that the problem of tractography is fundamentally ill-posed [Maier-Hein et al., 2017]. Estimating a global connectivity from local information about diffusion will inherently generate false positives and omit true positives. Even global tractography algorithms are not immune to this reality as they use the local voxel-wise information to connect segments together in order to generate plausible streamlines [Maier-Hein et al., 2017].

As for now, the majority of the tractography applications (clinical, surgical) are based on public software that implement iterative algorithms. As a consequence, this section is more specifically targeted towards algorithms using local models, more specifically iterative tractography algorithms (EUDX [Chamberland et al., 2014; Garyfallidis, 2013], DET, PROB [Jeurissen et al., 2017]).

## 2.2. TRACTOGRAPHY VERSUS UNDERLYING ANATOMY, WHAT IS AND WHAT ISN'T ?

### Bottleneck effect

The bottleneck effect is probably the most well-known limitation of dMRI tractography. When multiple fasciculi, or single fasciculus with multiple origins, converge toward a small region to then separate afterward, the observed diffusion signal will likely lead to a single orientation. During tractography, reconstruction at the bottleneck exit will have combinatorial possibilities and therefore can generate multiple bundles with invalid terminations as shown in Figure 2.2 and to some extent in Figure 2.4. When multiple pathways from various origins converge together to enter a region where they are all run parallel to diverge afterward, tractography cannot accurately “choose” the correct direction at the exit. Without explicit instructions, all combinations are equally valid, and such a configuration will always generate a large number of false positives. This phenomenon can be observed in the temporal stem, the corpus callosum or in the internal capsule. This was demonstrated in more detail in [Maier-Hein et al., 2017] by using a synthetic reconstruction of a subset of the long-range WM bundles.

Without advanced knowledge of the underlying anatomy, explicit rules, or priors, this situation cannot be solved. Simply stepping forward in an intricate vector field is not enough to avoid this pitfall. Bundle-specific tractography recently made an appearance in the literature, either using manually placed priors [Chamberland et al., 2017], using bundle templates [Rheault et al., 2019] or machine learning [Poulin et al., 2017; Wasserthal et al., 2018b]. All are a direct attempt to solve this problem and “teach” tractography how to untangle ambiguous signals by injecting information about the origin and the target of a streamline.

This effect is not unique to iterative tractography ; in a bottleneck region, even global tractography algorithms struggle to correctly assemble a streamline. There is still a combinatorial number of possibilities when it comes to connecting segments that go through a single-direction region.

## 2.2. TRACTOGRAPHY VERSUS UNDERLYING ANATOMY, WHAT IS AND WHAT ISN'T ?

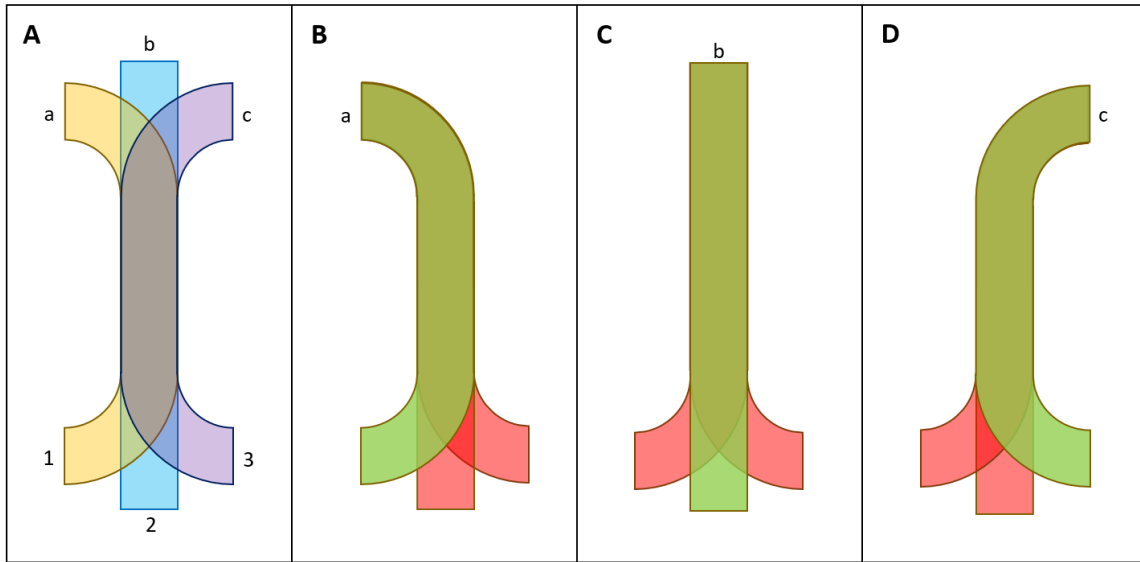


figure 2.2 – The bottleneck effect. Three (3) “*bundles*” with independent endpoints have an overlapping region (A), the bottleneck in shaded gray. In this region, all bundles run parallel and upon exit, the direction to take is unknown. In such a case, with 3 origins and 3 destinations, there are 9 potential outcomes (B-C-D), despite having only 3 valid configurations in the underlying architecture. This configuration will generate more invalid connections (IC) than valid connections (VC) as shown in [Maier-Hein et al., 2017]

### Wall effect

One of the lesser-known difficulties of tractography is encountered when a pathway crosses another at a low angle or in a kissing region. Kissing region occurs when bundles with different origins are curving and slightly overlap and then separate again as shown in Figure 2.3A. In this case, the reconstructed streamlines often look like they bounced off of each other. In regions where kissing pathways are expected, at the junction where they touch or overlap at a low angle, the reconstructed local model is not as sharp as it needs to be to disentangle both paths. When incoming streamlines take the curve and reach the overlapping region, they straighten their path due to the blurred single direction of the local model, until they exit the region. Instead of actually having a kissing region, the reconstruction appears to be separated by a “*wall*” as shown in Figure 2.3B and 2.4. In such a case, streamlines terminations

## 2.2. TRACTOGRAPHY VERSUS UNDERLYING ANATOMY, WHAT IS AND WHAT ISN'T ?

are considered valid, but the overall spatial extent is reduced. This can be observed around the centrum semiovale where the CC is restricted from entering the region as shown in Figure 2.11.

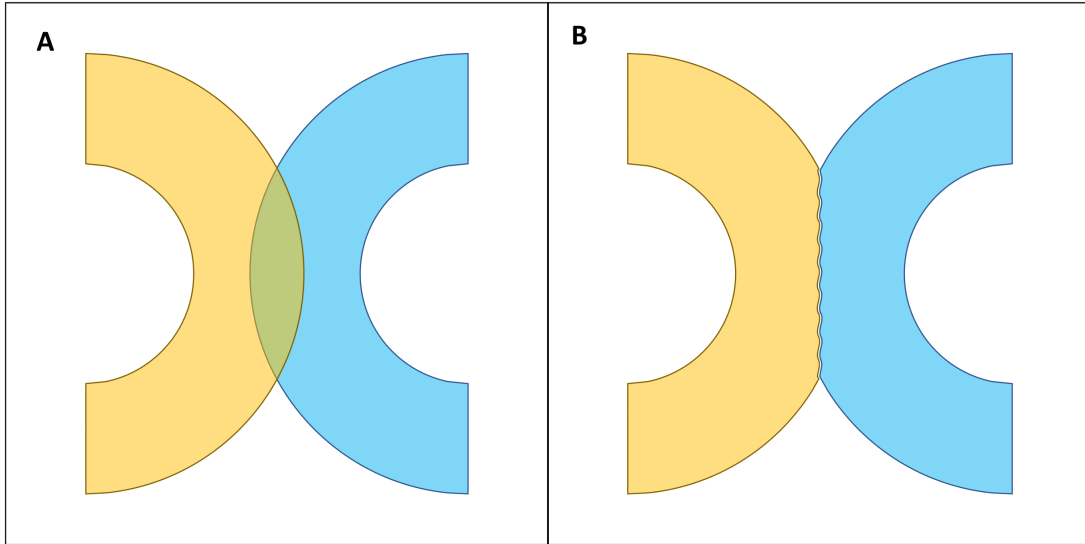


figure 2.3 – The wall effect. Two bending bundles with a small overlapping region (A). In this configuration, there is no invalid connection (IC), but the spatial extent of both bundles is limited at the “collision” region (B), which creates a wall effect and makes streamlines have a tendency to bounce off the other curving bundle.

## 2.2. TRACTOGRAPHY VERSUS UNDERLYING ANATOMY, WHAT IS AND WHAT ISN'T ?

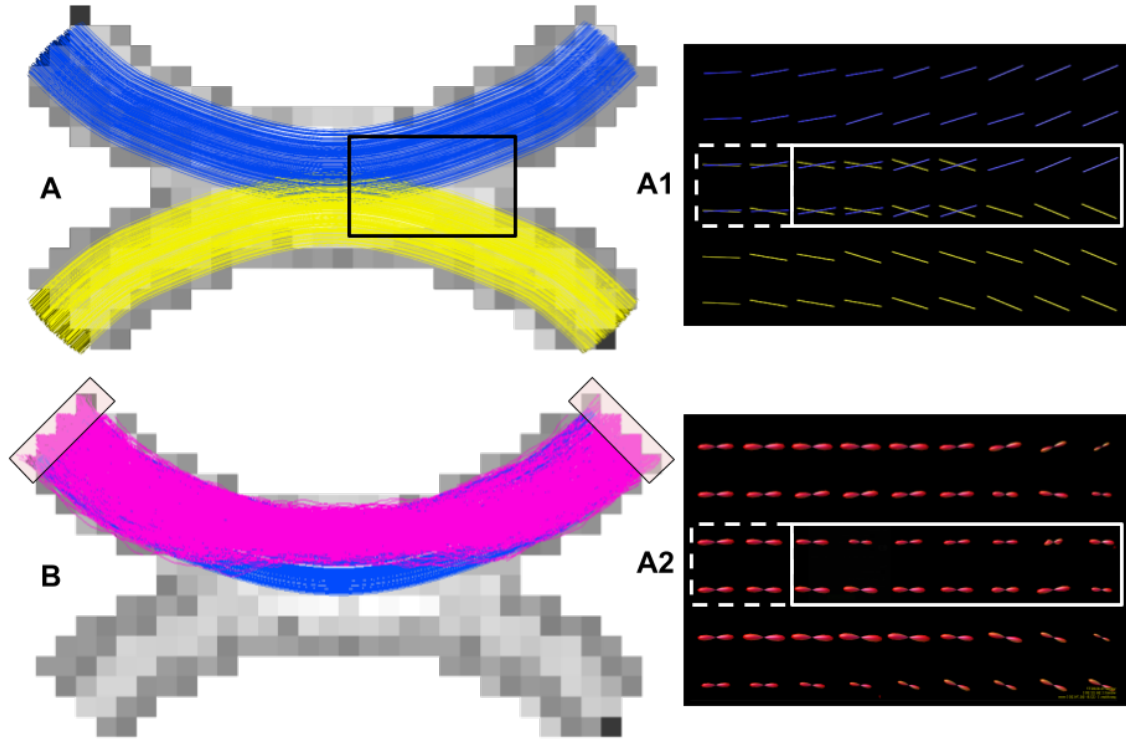


figure 2.4 – Representation of the wall effect using the open-source framework Fiberfox and an in-house geometry generator. In A, the “*ground truth*” composed of two bundles kissing. A1 represents the two separate vector-fields from the beginning of the overlap to the point where they run parallel. A2 shows the reconstructed fODF field, using CSD on the simulated data from Fiberfox. In both A1 and A2 the elongated white boxes show where the wall effect occurs, one could expect to see more than one fiber orientation if not for the wall effect. The dotted boxes show where the fODF are expected to contain a single orientation, this is due to the fact that the ground truth vector-fields are parallel at this location. In B, probabilistic tractography was performed, seeded from the boxes, on the fODF field. The pink streamlines show a clear decrease in volume when compared to the blue streamlines (ground truth).

### Narrow intersection effect

This effect is what happens when a streamline follows the wrong local direction in a crossing. When a streamline reaches a position where multiple directions are considered valid, it is possible that the algorithm simply picks the wrong one. The same can happen at the next step and so forth, with the errors accumulating until the

## 2.2. TRACTOGRAPHY VERSUS UNDERLYING ANATOMY, WHAT IS AND WHAT ISN'T ?

crossing region is over. Contrary to the previously described effect, this is not due to the fact that the local model is too simplistic when pathways overlap. In this case, the crossing can be correctly resolved in terms of local modelling, and it is simply due to the tractography algorithm picking the wrong direction when more than one choice is valid. Even at higher resolution or without noise, if the crossing angle is lower than the tractography curvature threshold, this effect will occur.

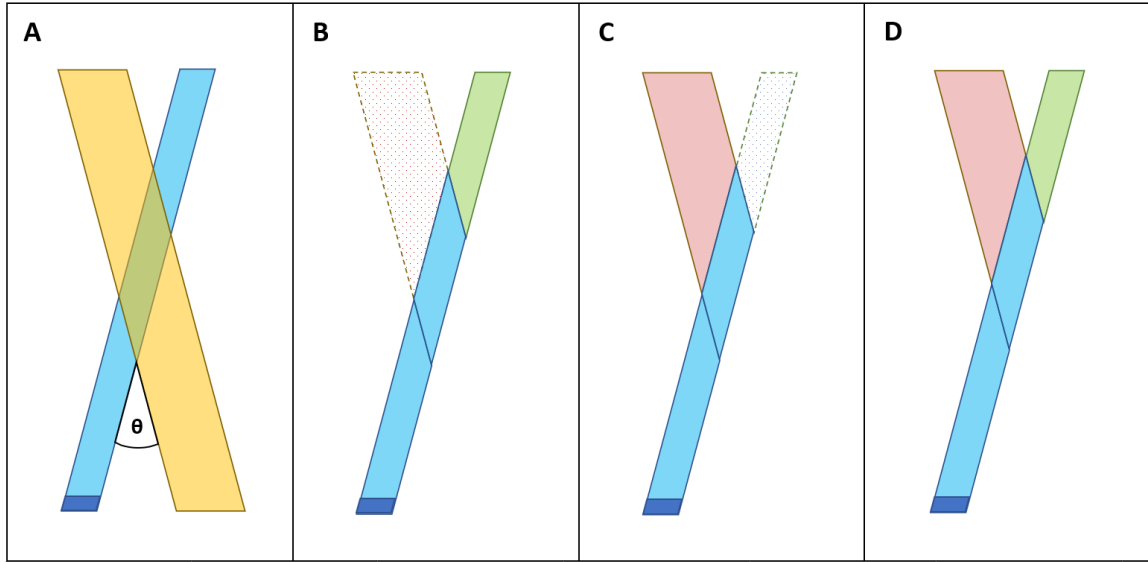


figure 2.5 – The narrow intersection effect. Two bundles crossing each other at a narrow angle  $\theta$  (A). In this scenario, the reconstructed local models correctly contain 2 independent lobes/peaks in the crossing region. The seeding region is shown using a dark blue box. If deterministic tractography (based on peaks, or fODF) is used, but the maximum curvature is limited to an angle below  $\theta$ , the bundle will be reconstructed correctly as shown in B. However, if the maximum curvature parameter is above the value of  $\theta$ , and the fODF value (amplitude) is higher for the bundle in yellow, deterministic tracking based on fODF will systematically take the wrong direction as shown in C. Probabilistic tractography based on fODF will reconstruct both paths as shown in C. The proportion of streamlines at both terminations will vary with the curvature parameters, fODF amplitude, the dimension and angle of the crossing.

This can be observed in the centrum semiovale, a well-known region often pointed out as a major pitfall for diffusion tensor imaging (DTI) [Descoteaux et al., 2009; Tournier et al., 2012]. Local models that support more than one direction per voxel

## 2.2. TRACTOGRAPHY VERSUS UNDERLYING ANATOMY, WHAT IS AND WHAT ISN'T ?

have their own limitations in this area. At high crossing angle, e.g above 45 degrees, this effect is not problematic. However, as the incoming angle decreases, over the course of a few voxels, some streamlines will go along the wrong direction, effectively “*switching pathway*” (see Figure 2.5). In the centrum semiovale region, the corpus callosum (CC) and PyT are fanning (lower density as the axons spread apart) and the AF is getting denser (stem region), giving an unfair advantage to the AF and limiting the two other. Within this crossing, multiple streamlines from the CC and PyT take a sharp turn in the anterior-posterior axis. In such crossing regions sharp turns are hard to make sense of, e.g is the turn representing the underlying anatomy of a single fiber population or does it represent two fiber populations joined by a single tractography misstep? Figure 2.6 shows sharp turns that are often hidden within well-known bundles. These can be difficult to interpret without knowledge of the anatomy.



## 2.2. TRACTOGRAPHY VERSUS UNDERLYING ANATOMY, WHAT IS AND WHAT ISN'T ?

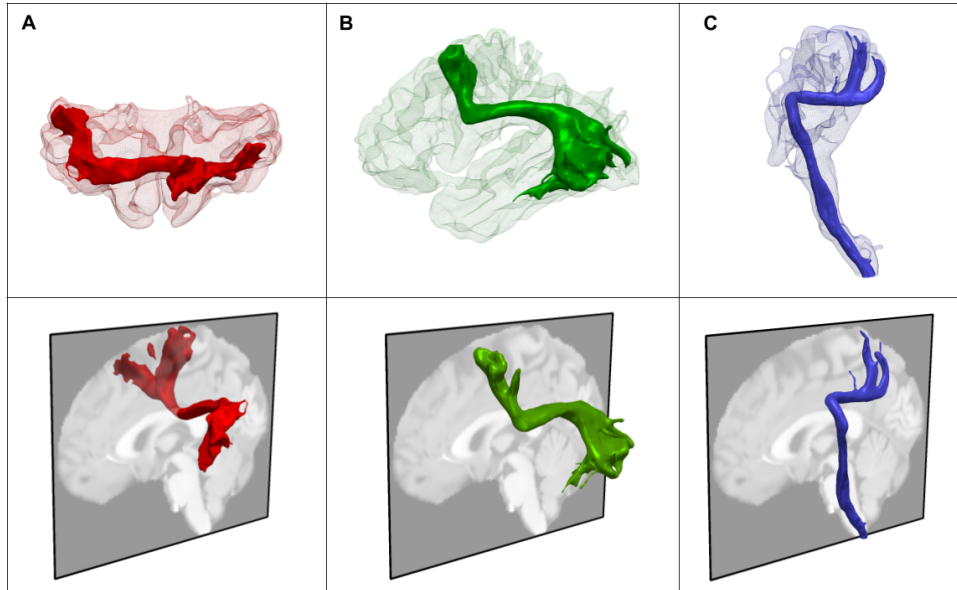


figure 2.6 – From left to right, the CC, AF and PyT. The transparent 3D surfaces represent the volume of the bundles as segmented. The solid 3D surfaces show the volume occupied by groups of streamlines that took an unexpected turn considering the overall shape of their respective bundle. In A, the left lateral portion of the CC seems to switch into the path of the AF. In B, a frontal portion of the AF rapidly turns into the gyri approximately where the frontal aslant tract (FAT) is located. In C, in the centrum semiovale the PyT takes a sharp turn in the antero-posterior axis following the path of the AF for a short distance to then return to its initial general orientation. It is difficult to know which of these turns are anatomically valid and which are artefacts; tractography alone is not enough to solve this issue, and anatomical priors/knowledge are required.

This can be observed even with global tractography algorithms, but the turns are usually less sharp. If the crossing happens at a low angle over a large region, streamlines can still “*switch pathway*” and, in such a case, invalid streamlines are even harder to distinguish from valid streamlines.

## 2.2. TRACTOGRAPHY VERSUS UNDERLYING ANATOMY, WHAT IS AND WHAT ISN'T ?

### 2.2.2 Spatial prior limitations

The termination criteria is one of the most influential decisions to take during a dMRI tractography project (along with the local model and the tracking type). Should the seeds be placed in the whole WM or just the WM/GM interface, and should the nuclei be included or not in the seeding region? Are the streamlines forced to finish at the cortex? How low does the projection streamlines have to go in the brainstem? Is tracking allowed in the nuclei? Can streamlines go into the cerebellum? Even when these questions can be answered with confidence, the origin and termination of streamlines is still subject to biases that cannot be easily solved. The following section describes limitations of tractography due to unknown rules on how tracking should behave in a particular region.

#### Gyral bias

One of the most well-known termination bias for dMRI tractography is the gyral bias [Reveley et al., 2015; Schilling et al., 2018a; Van Essen et al., 2014]. Due to the limitation in resolution and the often simplistic tracking methods used, the crowns are over-connected while the banks & fundi are under-connected. Another important consideration that amplifies the gyral bias is that gyri with complex convolutions make it difficult to even reach the crown. Any bundle or node in a connectome matrix related to such gyri is likely to be partially filled in terms of spatial extent and volume. Methods such as anatomically-constrained tractography (ACT) [Smith et al., 2012], particles-filtering tractography (PFT) [Girard et al., 2014] and more directly surface-enhanced tractography (SET) [St-Onge et al., 2018] attempt to reduce this bias, but the pattern is still observable despite the substantial improvement (see Figure 2.7). With the latest SET, as much as 30% of the cortical surface remains unreached and most banks and fundi are still empty [St-Onge et al., 2018], which presents a very serious open challenge for tractographers.

## 2.2. TRACTOGRAPHY VERSUS UNDERLYING ANATOMY, WHAT IS AND WHAT ISN'T ?

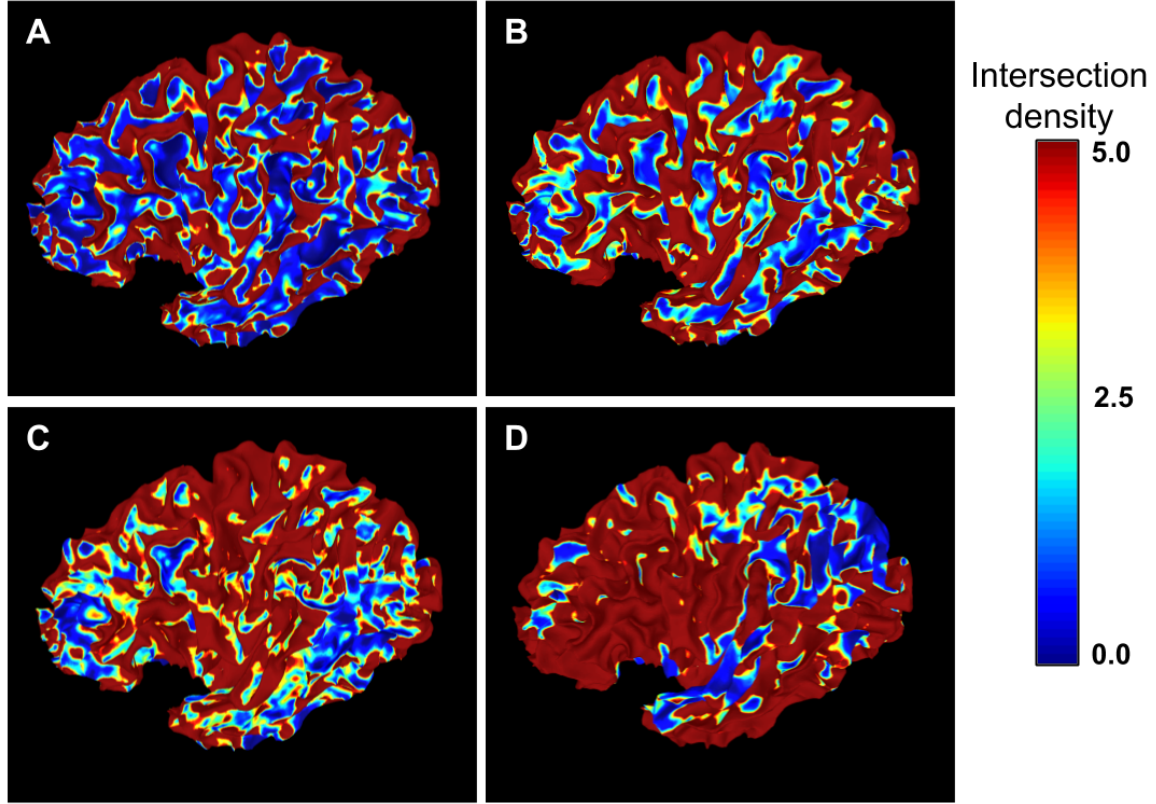


figure 2.7 – Cortical coverage of various tractography methods : Deterministic (A), Probabilistic (B), Particles-Filtering Probabilistic (C) and Surface-Enhanced Particles-Filtering Probabilistic (D) tractography. The coloring represents the number of streamlines intersecting with the surface per million streamlines. When the density per triangle is threshold at 2 intersections, the cortical coverage is 13.61%, 20.59%, 20.16% and 35.62% respectively.

### Deep nuclei bias

The deep nuclei provide another challenge for streamlines termination. Their complex structure often renders dMRI tractography uninterpretable within a nucleus [Zhang et al., 2018b], despite well-organized underlying WM pathways passing through the surrounding regions and along the surface. Short connections between nuclei are also very hard to interpret. Due to the simplicity of the local decisions during tractography, streamlines can (and often will) stop around the nuclei, blurring the line between the anatomically plausible streamlines and methodological artifacts.

## 2.2. TRACTOGRAPHY VERSUS UNDERLYING ANATOMY, WHAT IS AND WHAT ISN'T ?

This is due to partial volume effects and poorly defined termination criteria.

However, some of these streamlines could simply be passing by and running parallel for a short distance. More advanced algorithms using surfaces could prevent such terminations [St-Onge et al., 2018], but the lack of confirmation from anatomical validation makes it difficult to establish clear rules on what streamlines can and cannot do around the deep nuclei. Higher resolution will be a critical improvement to resolve the partial volume effect in these regions.

### Tissue classification limitations

Injecting more anatomical priors is likely the best way to reduce the inherent biases and pitfalls of tractography. However, good priors are quintessential to the process to guarantee that they do not introduce more problems. This is easier said than done, even for something as “*trivial*” as healthy tissue segmentation (WM, GM, CSF) from a structural T1w image.

For example, FSL-FAST [Woolrich et al., 2009] often classifies the deep nuclei as a mix of WM & GM (which enable multiple tractography algorithms to cross them), while ANTS-ATROPOS [Avants et al., 2011] classifies them as GM only and any streamline coming near them will stop. Each tissue classification algorithms will have slight variations, but consequences for tractography can be considerable as shown in Figure 2.8. Most segmentation tools inadvertently create a ring of GM around the ventricles due to partial volume effects between CSF and WM. Such a layer of GM wrongfully allows termination and leads to “*broken streamlines*”, which are perfectly valid in terms of shape, but abruptly stop in what should be WM [Girard et al., 2014; Smith et al., 2012; St-Onge et al., 2018]. Furthermore, the lack of differentiation between the cortical GM and the deep nuclei GM does not allow for spatially variant rules for tractography. Even if this was possible, such rules would need to be designed and confirmed by neuroanatomists.

## 2.2. TRACTOGRAPHY VERSUS UNDERLYING ANATOMY, WHAT IS AND WHAT ISN'T ?

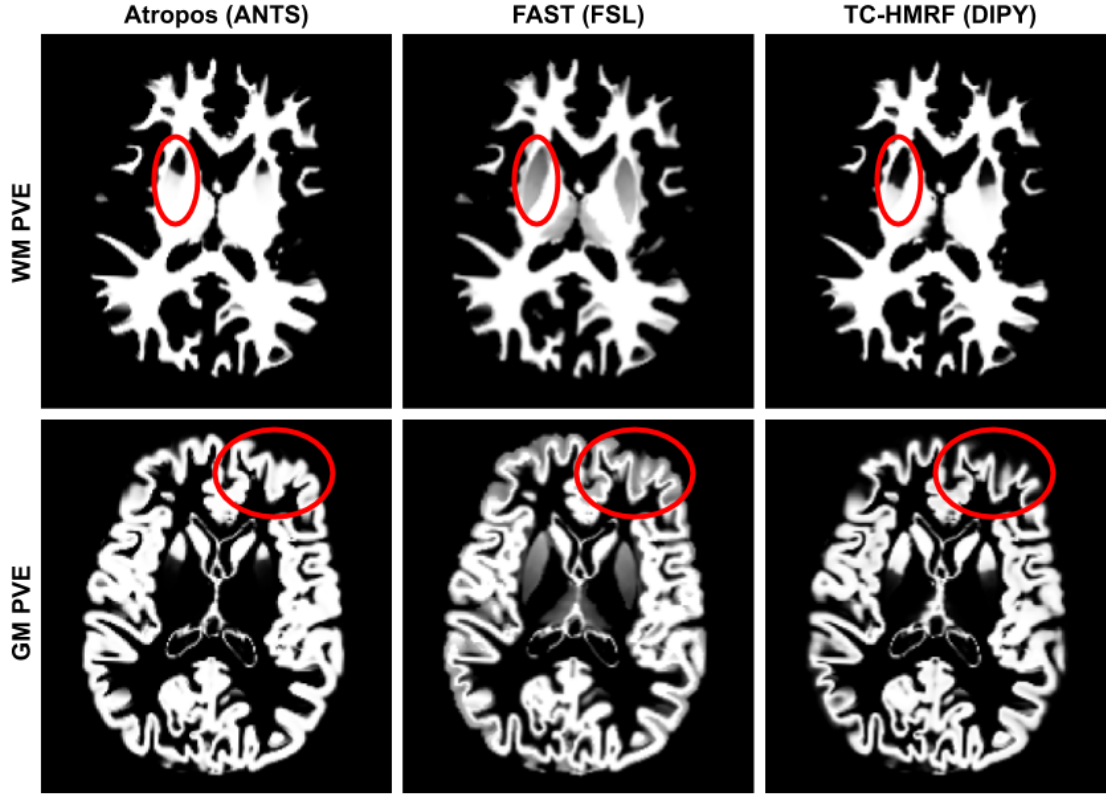


figure 2.8 – Resulting partial volume estimates (PVE) for GM and WM using 3 commonly used tools (ANTS, FSL and DIPY). The red circles in the WM maps highlight the difference in the putamen classification, each tool produces a variation of the size and certainty for the nuclei. Similarly, on the GM segmentation the cortical thickness is not the same and the probability is not as certain for FSL FAST. This mean that streamlines could come to an early stop at the cortex simply due to the tissue classification.

Proximity between GM regions and partial volume effects can reduce the area of WM “*passage*” or even completely block it. In the temporal stem, the GM “*leaks*” into WM, which sometimes prevents the inferior fronto-occipital fasciculus (IFOF) and uncinate fasciculus (UF) from being reconstructed completely [Hau et al., 2017]. Figure 2.9 shows the proximity of GM around the IFOF stem/bottleneck that could block the entire pathway for some subject or if tractography is performed with strict parameters.

## 2.2. TRACTOGRAPHY VERSUS UNDERLYING ANATOMY, WHAT IS AND WHAT ISN'T?

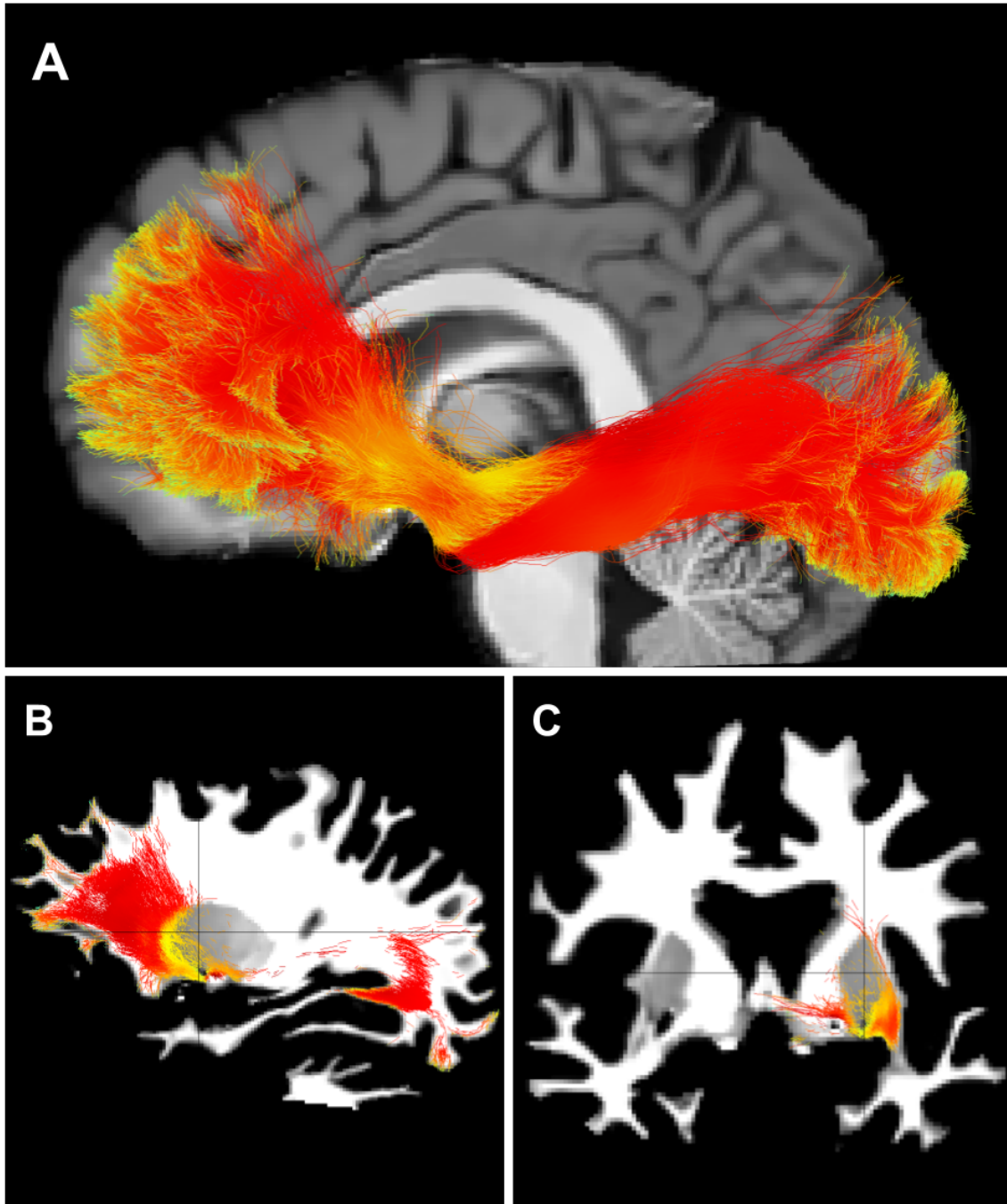


figure 2.9 – Reconstruction of the IFOF with color mapping from the underlying T1w values. Segments of streamlines colored in red are in the WM while the GM is in yellow and vary along the length of the bundle (A). The stem/bottleneck region is squeezed between the nuclei and the cortical GM (B & C). An algorithm that cannot traverse GM would not be able to reconstruct the IFOF.



## 2.2. TRACTOGRAPHY VERSUS UNDERLYING ANATOMY, WHAT IS AND WHAT ISN'T ?

Occlusion due to partial volume effect is still a common limitation of tractography algorithms using tissue classification as priors. In pathological cases with lesions apparent on the structural images, tissue segmentation is even more challenging or simply impossible and the only option is to use a fractional anisotropy (FA) threshold. However, if the FA threshold is too high (above 0.15) numerous crossing regions and most nuclei will not be included in the mask, and this could result into incomplete reconstruction of perfectly healthy bundles simply due to the nature of DTI and poor specificity of FA.

Limitations related to spatial and anatomical priors require not only an increase in resolution for T1w, but new algorithms based on more than one MRI contrast are likely necessary to fully solve these challenges. Algorithms dedicated to lesion segmentation will be necessary for each pathology [Atlason et al., 2019; Gryska et al., 2019; Subbanna et al., 2019]. Brain tumors, multiple sclerosis or stroke lesions all have different contrast signatures that make generalizations difficult.

### 2.2.3 Tractogram saturation

In principle, the volume of a segmented bundle should be invariant to the MRI type, acquisition scheme or tractography algorithm. That is, if we were able to “dissect” the entirety of the bundle and measure its volume, tractography or WM bundle segmentation would give us that number. However, in practice, tractography is not as straightforward, as seen in Figure 2.10. An unsaturated bundle means that not enough streamlines were generated to fully recover the expected spatial extent, which will lead to lots of downstream problems for quantification of tractography. DMRI tractography can be seen as a way to get a sample of a “theoretical” reconstruction, i.e a tractogram reconstruction is the realization of a stochastic process. This is why it is usually recommended to generate millions of streamlines (seeding) in order to obtain a representative tractogram reconstruction.

However, depending on the hypothesis being investigated or the processing power, it is not always possible to achieve a reconstruction with enough streamlines. Then, an obvious interrogation can be raised : how much streamlines is enough [Gauvin, 2016; Moyer et al., 2018] ? The answer is far more complex than what it seems at

## 2.2. TRACTOGRAPHY VERSUS UNDERLYING ANATOMY, WHAT IS AND WHAT ISN'T ?

first glance. The consequences of not generating enough streamline depends on the intended use, meaning that it will vary with the hypothesis and interpretation. This is often referred to as “*bundle saturation*”, where the volume of a reconstructed bundle does not change even if more streamlines are generated. Figure 2.10 shows that some bundles or algorithms stabilize faster than others.

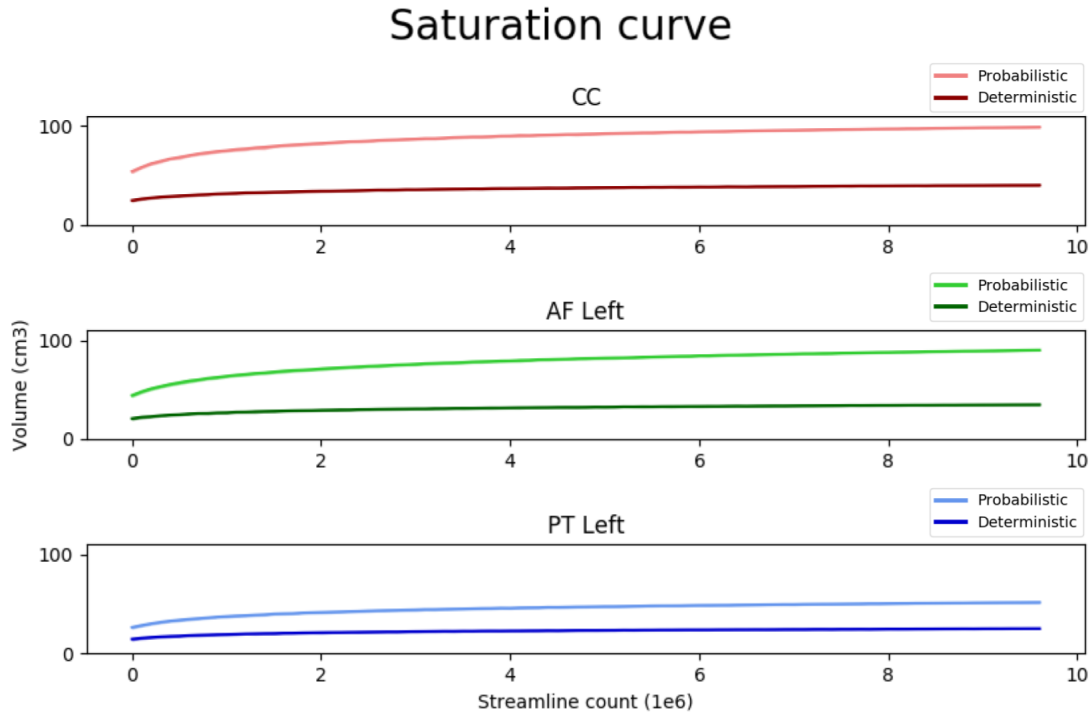


figure 2.10 – Saturation curve of 3 bundles, the CC (red), AF (green) and PyT (blue) using deterministic (darker) and probabilistic (lighter) tractography. On the X-axis, the number of generated streamlines (above 40mm) in the tractogram up to 10 million streamlines. On the Y-axis, the volume in  $cm^3$  of the segmented bundle of interest. Deterministic tractography reconstructs a smaller volume and converges/stabilizes faster than probabilistic tractography. It is important to note that the scaling of X-axis is compressed for visualization, giving the impression of ever-increasing volume. For example, the probabilistic CC at 100k and 200k streamlines has  $54.1cm^3$  and  $58.2cm^3$  respectively, and at 9.9M and 10M streamlines the volume is at  $98.9cm^3$  and  $99.0cm^3$  respectively.



## 2.2. TRACTOGRAPHY VERSUS UNDERLYING ANATOMY, WHAT IS AND WHAT ISN'T ?

### Seeding strategies

Seeding strategies are crucial when interpreting generated tractograms. It is now commonly known [Girard et al., 2014; Smith et al., 2012] that seeding from the WM voxels allows for an easier reconstruction of large volume bundles than smaller ones. Indeed, as they occupy more space, they will receive more seeds and more streamlines will participate in their reconstruction. Hence, they will tend to be over-represented in the connectome. Another example is the node weight in connectome matrices, deterministic algorithms will stabilize faster than probabilistic algorithms, but not all nodes will stabilize at the same time [Yeh et al., 2018a].

Seeding from the WM/GM interface (or more recently surfaces) will reduce this bias [Girard et al., 2014; St-Onge et al., 2018]. As opposed to being biased towards bundles with large volume, streamlines distribution is linked to the area at the WM/GM interface. These techniques will make the small hard-to-track bundles that cover only a small cortical area invisible/untrackable/unobserved or harder to track, such as the fornix [Rheault et al., 2018a]

Overall, poor seeding can easily lead to unobserved pathways, while seeding too much can lead to the observation a large number of uninterpretable pathways. It is also important to mention that global tractography is not immune to this problem. Factors such as segment length, voxel size, and number of seed initialization will influence the number of streamlines needed to fully saturate a tractogram.

### 2.2.4 Obstacles along the path

Limitations due to local models are the main source of error for dMRI tractography, but it is important to consider them on a larger scale. Decisions made at the voxel level are hard to inspect, but their impact at the scale of the whole brain is more easily understood. Overall, the effects described in Section 2.2.1 can be seen as “*difficult regions*” or even “*obstacles*”. When looking at a reconstructed bundle and the amount of streamlines in it, it is crucial to consider the number of such obstacles in its way.

## 2.2. TRACTOGRAPHY VERSUS UNDERLYING ANATOMY, WHAT IS AND WHAT ISN'T ?

One of the reasons why not all bundles are as easily reconstructed as other is that each bundle has its own obstacles along its path. Bundles with a constantly curving path passing through crossing regions will be harder to reconstruct than straight bundles. This is one of the reasons why the fanning of the PyT is harder to reconstruct than the “*core*”. This is not even mentioning obstacles related to lesions (segmentation, termination criteria, etc.), bundles strongly affected by a pathology could be badly reconstructed simply due to poor tissue segmentation and not because of the diffusion signal and tractography themselves. When judging the results of a tractography algorithm, it is important to consider all these obstacles in order to evaluate if the reconstruction truly represents the diffusion signal and the underlying anatomy or if it is simply limited by the processing itself. As presented in Figure 2.11, the resulting spatial distribution of streamlines following a bottleneck region can hardly be attributed to the signal or the anatomy, but rather to algorithmic limitations.

Furthermore, the longer a bundle is, the higher the probability of encountering obstacles will be. In practice, this means that more “*good*” local decisions are required to avoid mistakes. In the case of the PyT, the streamlines all have to navigate through areas of partial volume effects in the brainstem, traverse the internal capsule surrounded by nuclei, and finally follow the complex fanning structure into a large area of the cortex. As a result, streamline density in a bundle is often strongly linked to distance because of these encounters with difficult regions [Rheault et al., 2019; Sotiropoulos et Zalesky, 2019]. These biases are strictly due to processing, reconstruction and algorithmic choices and not related to the true underlying anatomy.

## 2.2. TRACTOGRAPHY VERSUS UNDERLYING ANATOMY, WHAT IS AND WHAT ISN'T ?

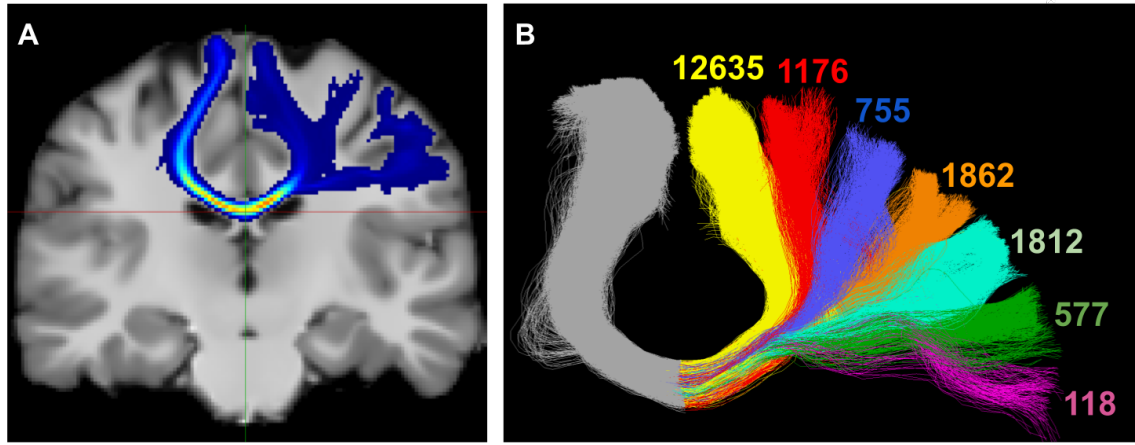


figure 2.11 – Obstacles along the path. Cross section of a CC density map (A) and the streamline count distribution along the precentral gyrus (B) using probabilistic tractography. In this scenario, the Corpus Callosum bottleneck makes it difficult to disentangle the target of a streamline and the distribution cannot easily be linked to the underlying anatomy. Then, at the boundary of the centrum semiovale, it is likely that a wall effect prevents streamlines from entering the crossing, causing the large difference between the first two regions (yellow & red). Finally, the crossing itself is another difficult area to navigate, and since this is probabilistic tractography output, it is likely that streamlines “switched pathway”. All these obstacles will affect the distribution of streamlines, which is unrelated to the underlying anatomy.

Post-processing algorithms such as SIFT, LIFE, and COMMIT [Daducci et al., 2015; Pestilli et al., 2014; Smith et al., 2015a] attempt to fix this disparity in density that does not represent the local signal. By assigning a weight to each streamline to explain the observed diffusion signal, streamlines can be discarded if they do not provide a significant contribution. These algorithms necessitate tens of millions of streamlines to guarantee that reconstructed bundles reach their full spatial extent, which leads to the emergence of somewhat dense false positives. Dense anatomically implausible grouping of streamlines can sometime explain the signal better than a few plausible streamlines. Injecting more anatomical priors into the tractography algorithms themselves is probably necessary in order to decrease the amount of invalid streamlines, which will make them more easily identifiable [Poulin et al., 2019; Schiavi et al., 2019].

## 2.3 Individual pathways and connectomes, how to (not) interpret it ?

Once a tractogram has been generated in the best possible way, one could think the hardest part is over. This could not be more false, because a tractogram is useless without a way to analyze what is inside. Clinically oriented research typically uses one of the two most widespread approaches : bundle segmentation or connectomic. If not handled with care, these approaches aggravate the previously presented biases and pitfalls, worsen misconceptions and lead to erroneous interpretations and conclusions. This section quickly summarizes how such analysis can go wrong and which factors need to be carefully thought out before starting a project involving one of these approaches. It is important to specify that tractograms based on DTI, HARDI tractography, or global tractography will be subject to these biases, and that none is immune.

### 2.3.1 Bundle segmentation

#### Manual approach

Despite being heavily time consuming, manual segmentation is still very much used by neuroanatomy experts. By carefully delineating regions of interest (ROIs) on which to apply sets of rules, a rater (the person doing the segmentation) isolates a bundle of interest (BOI). Evidently, all previously mentioned biases and pitfalls will influence how the tractogram is reconstructed and any bundles dissected from it. However, the act of performing the segmentation comes with its own biases.

The first bias is a validation bias from the raters, when “*wrong-looking*” streamlines are more easily accepted in hard-to-track bundles in order to fit the dissector’s priors. Bundles that are harder to reconstruct often have a poor spatial extent if the segmentation is too strict. Thus, the ROIs are typically made larger in order to allow more streamlines to be filtered, even if their position and shape does not entirely respect the anatomical definition. For easier-to-track, long-range bundles, a dilemma appears between having a liberal or conservative set of rules. A more liberal approach comes with more manual cleaning, which is time consuming. On the other hand, a

### 2.3. INDIVIDUAL PATHWAYS AND CONNECTOMES, HOW TO (NOT) INTERPRET IT?

conservative approach will reduce outliers at the price of potentially discarding anatomically plausible streamlines. Slightly increasing the size of ROIs can drastically change the resulting number of streamlines and the volume they occupy, as shown in Figure 2.12.

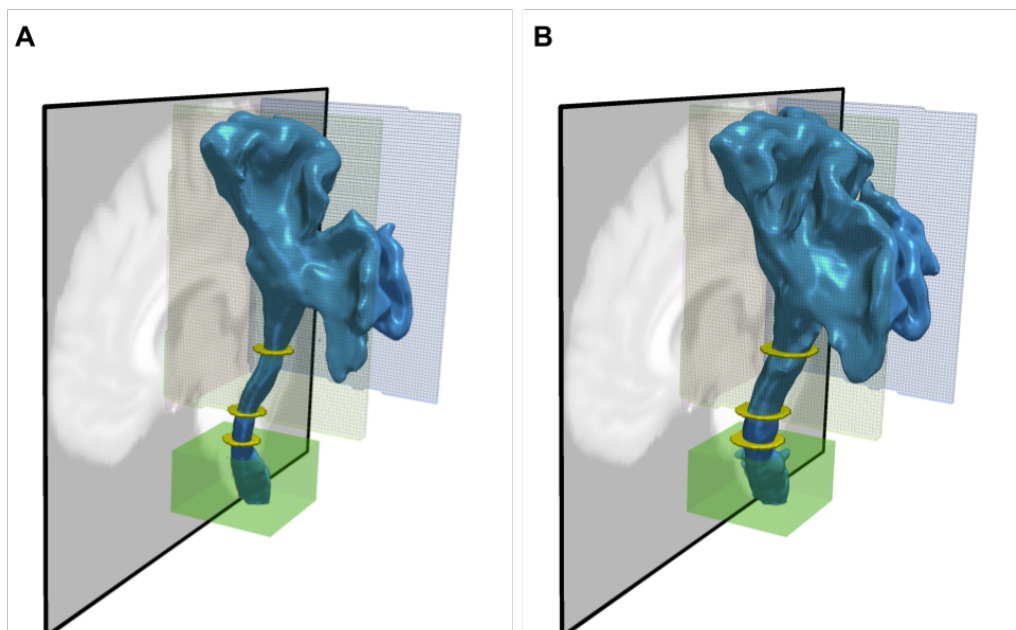


figure 2.12 – Two segmentations of the PyT, one is stricter (A) and the other is more liberal (B). Both were performed using rules enforcing not to go over the midline or more anterior to the precentral gyrus and more posterior to the postcentral gyrus (3 transparent planes) and down the brainstem (box). 3 points of passage (internal capsule, midbrain and medulla oblongata) were used to isolate the path. The conservative version used planar disks of 5mm radius and the liberal version used planar disks of 7.5mm radius (disks were enlarged for visualization). The spatial extent of the fanning increases to completely cover both gyri and the core of the bundle becomes thicker in the liberal segmentation. The volume of the conservative version (A) is  $36.2\text{cm}^3$  and the liberal version (B) reaches  $58.0\text{cm}^3$ .

The second bias is due to intra-rater and inter-rater variability [Gwet, 2012]. Since dMRI tractography is a virtual method, experimental variability can be assessed. The same anatomical definition, on the same tractogram, from the same rater, at two different time-points, will inevitably produce slightly different results. The same

## 2.3. INDIVIDUAL PATHWAYS AND CONNECTOMES, HOW TO (NOT) INTERPRET IT ?

thing happens if the segmentation is done by other raters [Rheault et al., 2020]. This bias could become a major pitfall if the set of rules is not clear enough or prone to error, increasing the variability too much. Observations and interpretations based on a single segmentation could change if performed by a different rater, especially if this variability is not accounted for in the statistical analysis. In such a case, conclusions could be erroneous.

### **Automatic approach**

While automatic segmentation results do not have intra-rater and inter-rater variability, they however rely on algorithms that are often not free of their own problems. It is important to consider that most automatic methods claiming to reconstruct anatomically relevant bundles rely on some sort of registration algorithm, which can be strongly affected by datasets containing pathological or aging brains. Such datasets always present a challenge, and using automatic methods designed on healthy datasets can have unexpected results.

Another common problem with automatic methods is that they heavily rely on pre-determined definitions of bundles. Such definitions are sometimes not fully agreed upon and could even be considered wrong (mentioned in section 2.3.2). However, the ease of use of an automatic method on large datasets makes it interesting despite the anatomically dubious definitions. Without even realizing it, newly developed methods participate to the confusion. Sometimes developed by researchers with no background in anatomy, new methods are presented purely for algorithmic consideration, and can add “noise” to the literature when it comes to bundle segmentation/definition. If the lack of consensus in anatomical definition is addressed in the future, any automatic method that would follow the agreed upon definition would be a very interesting asset for the field of dMRI tractography. In the meantime, when using an automatic segmentation method, the resulting bundle of interest should be considered with caution.

### 2.3. INDIVIDUAL PATHWAYS AND CONNECTOMES, HOW TO (NOT) INTERPRET IT ?

This section will present a limited overview of known methods, where algorithms were not chosen for their popularity but rather their differences with each other, which highlight various considerations.

**TractQuerier (point-based) :** [Wassermann et al., 2016] This automated tool has a very simple but powerful concept, which is to define a set of rules in a query language based on an atlas of ROIs. However, the complexity of the query increases rapidly for several bundles in order to truly capture the proper shape. Defining endpoints is easy, but most atlases are limited in terms of WM parcellation, which makes the passage points hard to define and often lead to outliers. Capacity to generalize to pathological or aging brains will depend on the registration algorithms and the choice of atlas. If the tractogram was “free” of artifacts and spurious streamlines, TractQuerier is, in theory, the tool of choice.

**Tracula (signal- & position-based) :** [Yendiki et al., 2011] This tool is available in the Freesurfer package uses prior distributions on the surrounding anatomical structures are learned from the Freesurfer cortical parcellation to constrain the tractography solutions for each pathway. The underlying method is very similar to global tractography (where the entire path between regions is estimated at once), but output volumetric probability (voxel-wise) instead of streamlines. The spatial relationship between bundles makes the reconstruction dependent on what is around it, which is an important concept in neuroanatomy. As opposed to TractQuerier, the pathways are pre-determined ; this is because the method ultimately relies on manual labelling in a training set. The generalization to pathological and aging brains has been investigated and have been shown to be robust [Storsve et al., 2016].

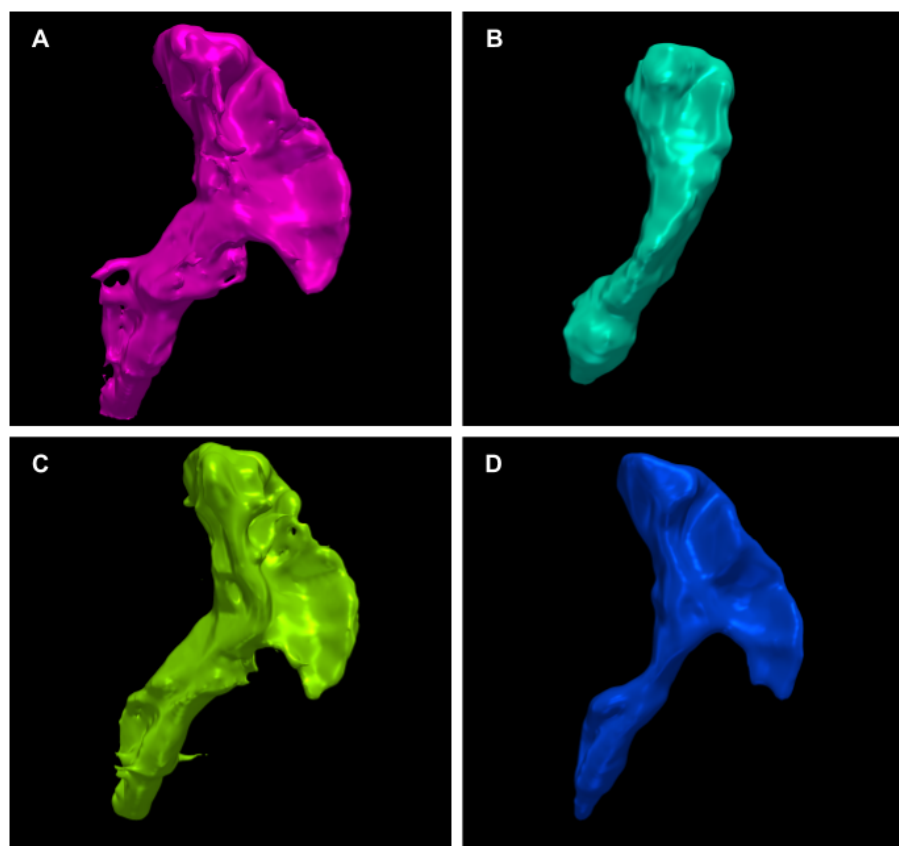
### 2.3. INDIVIDUAL PATHWAYS AND CONNECTOMES, HOW TO (NOT) INTERPRET IT ?

**Recobundles (streamline shape-based) :** [[Garyfallidis et al., 2018](#)] This tool from the Dipy library aims to segment bundles based on their shape similarity to model bundles, part of a dictionary (e.g build from expert dissections). Recobundles is less reliant on registration than most tools, but heavily dependent on the model bundles and the fine-tuning of parameters. By according more importance to the general shape rather than exact anatomical rules and terminations, this method was shown to be robust to various pathologies.

**Tractseg (peaks- & position-based) :** [[Wasserthal et al., 2018a](#)] This tool is based on convolutional neural networks that perform bundle segmentation on the field of FODF peaks (principal directions). It does not create tractography per se, but it was recently used in conjunction with tract orientation mapping to achieve bundle-specific tractography [[Wasserthal et al., 2018b](#)]. These two methods are among the most promising approaches as they could learn to navigate the field of directions using complex anatomical priors that vary according to their position. However, generalization to unseen datasets and data harmonization is a major difficulty for machine learning methods [[Poulin et al., 2019](#)]. It also heavily relies on the choice of anatomical definitions for the manual labelling used in its training data. Model bundles used for training were based on TractQuerier, this mean as it incorporates advantages and downsides of its anatomical definition.



### 2.3. INDIVIDUAL PATHWAYS AND CONNECTOMES, HOW TO (NOT) INTERPRET IT?



	TractQuerier	Tracula	Recobundles	TractSeg
TractQuerier (A)	1.00	0.543	0.491	0.594
Tracula (B)	0.543	1.00	0.448	0.661
Recobundles (C)	0.491	0.448	1.00	0.463
TractSeg (D)	0.594	0.661	0.463	1.00

figure 2.13 – Each box shows an automatic method that segmented a bundle of interest (CST) from the same input data (HCP). The methods do not follow the same anatomical definition and therefore do not reconstruct the same spatial extent. In order, TractQuerier (A), Tracula (B), Recobundles (C) and TractSeg (D). TractQuerier requires streamlines as input, standard query using the Desikan-Killiany atlas and a probabilistic tractogram were used. Tracula requires the raw DWI and the complete Freesurfer output to estimate the local model (bedpostx) and their priors. Recobundles requires streamlines as input, standard models provided in Dipy [Garyfallidis et al., 2014; Yeh et al., 2018b] and probabilistic tractogram were used (same as TractQuerier). TractSeg requires the raw DWI as input (same as Tracula).

## 2.3. INDIVIDUAL PATHWAYS AND CONNECTOMES, HOW TO (NOT) INTERPRET IT?

### 2.3.2 Debate over anatomical definition of pathways

Cortex electrical stimulation during surgery to guide resection [Suess et al., 2006; Szelényi et al., 2010], strokes location to predict deficits [Mazzocchi et Vignolo, 1979; ROBINSON et al., 1984], and deep brain stimulation for Parkinson [Benabid, 2003] or Epilepsy [Sprengers et al., 2017] are examples that seem to indicate that our understanding of the human brain is quite advanced. However, these examples are in fact much more about knowing the cortex and deep nuclei and not about the connectivity between various brain regions. Even with the seminal work of [Dejerine et Dejerine-Klumpke, 1895] dating to more than a century ago, the current state of consensus on anatomical definitions of WM pathways is far from perfect. Even when knowing that two regions are indeed connected, the exact path is not as clearly defined as some would expect.

Slight variations in terminology across laboratories can sometime explain differences of pathways, such as the case of the pyramidal tract (PyT). This is a good example of a well-known pathway with a range of possible, almost equivalent, definitions. First described by [Déjerine et Dejerine-Klumpke, 1901], it is now defined as starting from the cortex, going through the internal capsule, then through the anterior region of the brainstem and finally to the medulla oblongata within the pyramids [Chenot et al., 2019]. It is generally considered to encompass both the corticospinal tract (CST) and the corticobulbar tract (CBT). Since the PyT is composed of motor neurons, some definitions require terminations in the precentral and postcentral gyri, while others include any neurons that originate at the cortex and continue past the lower brainstem [Ebeling et Reulen, 1992]. One can easily see multiple ways to describe the almost exact same pathway without having an exact agreement on the anatomical definition.

For example, neuroanatomists using dMRI tractography describing the PyT could define it as “Cortex to lower brainstem” while others define it as “Cortex to medulla oblongata” which is very similar in essence. However, the CBT does not reach as low as the medulla oblongata, and that would mean that the generally agreed upon definition (that the PyT encompasses the CST and CBT) is not respected [Chenot

## 2.3. INDIVIDUAL PATHWAYS AND CONNECTOMES, HOW TO (NOT) INTERPRET IT?

et al., 2019]. Such discrepancies across experts in neuroanatomy are common [For-  
kel et al., 2014; Mandonnet et al., 2018; Panesar et Fernandez-Miranda, 2019]; even  
though an agreement may have been reached on a general anatomical definition, slight  
differences can lead to major deviations in the final results (and may cause unneces-  
sary confusion). Only some specific WM pathways can be considered controversial,  
but even the most well defined pathways can be subject to variation in anatomical  
definition due to “noise” in the tens of publications that relay information about its  
characteristics. While none of these descriptions are perfect, some are better than  
others.

This problem is by no means specific to the field of dMRI tractography. Even  
brain structures as well known as the hippocampus are not as well defined as one  
would think. Historically, the description of the structure has varied a lot (position,  
shape, extent), but protocols on how to perform a 3D virtual segmentation have  
fluctuated even more. A review of the existing (and commonly used) hippocampus  
segmentation protocols showed that the agreement across protocols was far from per-  
fect [Boccardi et al., 2011]. There was no clear consensus in neuroimaging/radiology  
on the anatomical definition of the hippocampus. Steps were taken to create a harmo-  
nized protocol that would minimize the chance of error, maximize reproducibility and  
reduce confusion [Boccardi et al., 2015]. Such protocol allows researchers to "pull/ag-  
gregate data together" across studies to perform meta-analysis and facilitate literature  
reviews [Frisoni et al., 2015]. On the long term, the very same steps will need to be  
taken for every major WM pathway of the human brain [Rheault et al., 2020]. In  
the meantime, caution is advised when tractograms are segmented using unvalidated  
anatomical definitions, especially when directly linking results to underlying anatomy  
or claiming to solve or understand cognitive processes.

### 2.3.3 Connectomics

All the tractography biases and pitfalls previously mentioned will influence the  
tractogram, which is the key element of a connectome matrix reconstruction. However,  
when it comes to building a connectome matrix, the tractogram is only half of the  
equation, and the remaining factors, such as i) choice of atlas, ii) size of parcels, iii)

### 2.3. INDIVIDUAL PATHWAYS AND CONNECTOMES, HOW TO (NOT) INTERPRET IT ?

parcel borders, and iv) weighted or binary, are just as important [Qi et al., 2015]. In this section we will focus not on dMRI tractography, but on the factors that must be considered after tractogram generation and how they can affect results, interpretations and conclusions based on connectome matrix.

#### Choice of atlas

Brain parcellation is far from being a solved problem. A wide variety of atlases are now readily available, each using a different approach to subdivide the brain morphology [Desikan et al., 2006] (Figure 2.14A), cytoarchitecture [Eickhoff et al., 2005], functional connectivity [Thomas Yeo et al., 2011; Tzourio-Mazoyer et al., 2002] (Figure 2.14B-C), and structural connectivity [Glasser et al., 2016; Tittgemeyer et al., 2018], or using mixed-approaches combining two or more factors [Fan et al., 2016] only to name a few. Numerous cortical atlases exist, but exclude the nuclei, which act as a relay system and are crucial in various cognitive functions, or the brainstem, which is crucial to various roles in the regulation of basic functions and in the motor and sensory system. Excluding or including such regions in the connectome matrix should be a careful decision. The intended goal is key to reach an appropriate decision for the atlas [Qi et al., 2015; Wei et al., 2017].

#### Size of parcels

Largely the most influential non-anatomical choice for connectome building [Sotiropoulos et Zalesky, 2019; Yeh et al., 2018a]. In the case of specific atlas choice, parcels size cannot be manipulated, however it is important to consider this factor when "shopping" for an atlas. Large parcellations will likely be more robust to inter-subject variability, but have a lesser ability to discriminate between observed variables (e.g cognitive functions). Increasing the parcel-count can increase specificity, but can be problematic for statistical analysis (e.g number or statistical tests) [Sotiropoulos et Zalesky, 2019; Thirion et al., 2014]. Parcellations are sometimes available at different resolutions, increasing at each layer the number of parcels to more clearly define known anatomical landmarks [Fan et al., 2016; Ghaziri et al., 2018; Hagmann et al., 2008]. As mentioned in Section 2.2.2 tractography has a hard time reaching the

### 2.3. INDIVIDUAL PATHWAYS AND CONNECTOMES, HOW TO (NOT) INTERPRET IT ?

cortex. A parcellation thick enough (reaching into WM) near the cortex is important in order to allow streamlines to reach the parcellation. Thin cortical bands that do not even reach into WM often discard a lot of streamlines simply because they cannot actually reach the regions of interest by a couple of millimeters.

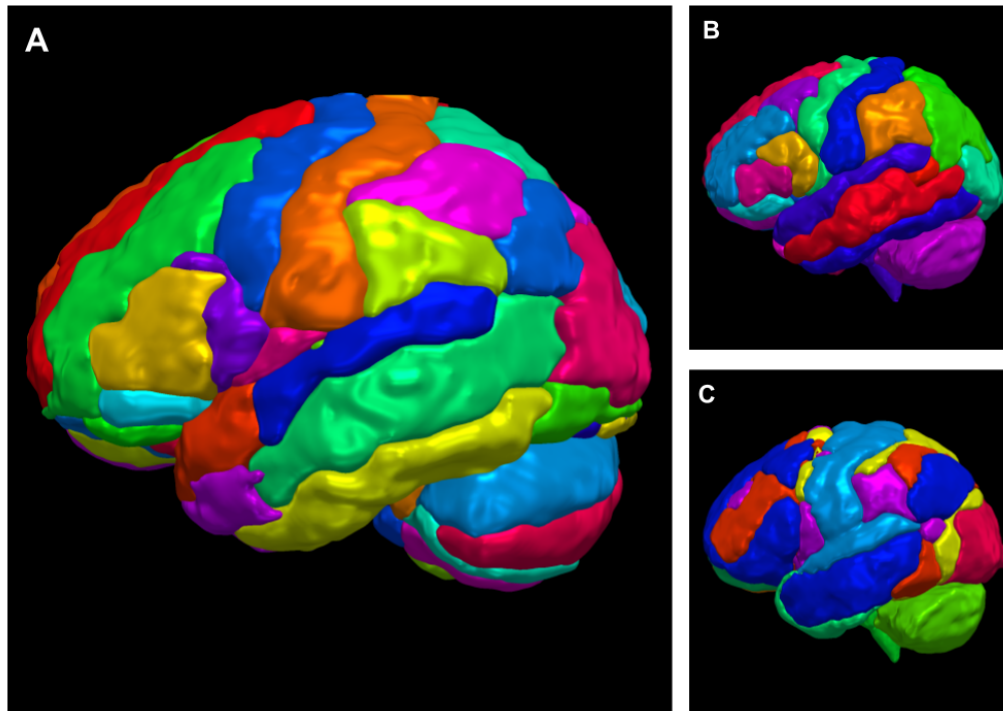


figure 2.14 – Parcellation of the cortical GM, brainstem, cerebellum and deep nuclei come in a wide variety (Desikan-Kikenny (Freesurfer) in A [Desikan et al., 2006], Automated anatomical labelling atlas (AAL) in B [Tzourio-Mazoyer et al., 2002] and the 17 functional network of [Thomas Yeo et al., 2011]. Subdivisions of various size and can vary in precision spatially (more subdivisions in one region in particular) will have consequences on the observed results and interpretations [Qi et al., 2015; Sotiropoulos et Zalesky, 2019; Wei et al., 2017].

#### Parcel borders

Defining the border of parcels is a harder task than one could suspect [Yeh et al., 2019], not only in terms of algorithmic or processing but even in terms of neuroanatomy. Frontiers between brain regions are hard to define clearly across a population.

## 2.4. DISCUSSION AND CONCLUSION

Anatomical variability is typical (even in a closely related group) and some cortical regions cannot be clearly defined. Streamline terminations are not necessarily valid even if their entire path is valid, and as such the discrete nature of a streamline should not be confused with certainty. Increasing the number of parcels will likely cause more uncertainty at the frontier between each pair of parcels. To address the difficulty to assign streamline ending at the frontiers of parcels, the concept of “*fuzzy boundaries*” has been proposed [Sotiropoulos et Zalesky, 2019].

### Weighted or Binary

Last but not least, the decision to weight the connectome matrix needs to be made. Should it be weighted by streamline count (most popular approach), FA, volume, length or something more complex [Colon-Perez et al., 2015; Sotiropoulos et Zalesky, 2019], should it be binarized [Civier et al., 2019; Sotiropoulos et Zalesky, 2019]? All options are available, however it is important to mention that if the goal is to quantify the edges (streamlines) in order to get neuroanatomical insight, it should be interpreted with extreme caution. Publications sometimes used connectome weighted by FA to then claim FA was a measure of WM integrity (it is not) and that overall FA-weighted connectome was a way to evaluate brain efficiency (it is not). Such over-interpretation of diffusion measurements and tractography should be avoided [Jones et al., 2013]. Interpretation of connectome is more than tricky, it is however an useful tool for prediction using correlation, without trying to interpret the underlying anatomy. A connectome matrix is an effective/useful data-driven summary of a whole brain tractogram.

## 2.4 Discussion and Conclusion

The purpose of this review is to put into perspective the various limitations for the field of dMRI tractography in general. Limitations of various algorithms and of our knowledge of neuroanatomy must be recognized explicitly in order to avoid potential shortcomings in the future. Acknowledging these challenges is a great starting point for PhD students looking for projects or to justify their new solution or algorithm. This review should be seen as constructive criticism for dMRI tractography, aiming

## 2.4. DISCUSSION AND CONCLUSION

to increase understanding of the various limitations of all tractography methods and raising awareness. It is crucial to understand that all tractography algorithms are prone to all presented biases (to various severity). Tractography performed on high quality datasets, such as the Human Connectome Project (HCP) [Glasser et al., 2013; Van Essen et al., 2013], is not immune to these biases or pitfalls. Poor understanding of an algorithm should not be an excuse for misinterpretation. Most biases can be acknowledged without truly affecting the discussion or conclusion of a publication, as long as they are mentioned and well-explained. This review was only a brief summary of often previously mentioned limitations [Jbabdi et Johansen-Berg, 2011; Jones et al., 2013; Maier-Hein et al., 2017; O'Donnell et Pasternak, 2015; Sotiropoulos et Zalesky, 2019], understanding the limitations of dMRI tractography is as important as understanding its strengths.

Reporting negative results, publishing reproduction studies and promoting open software are all necessary steps to increase transparency of results. As of now, each methodology has its own hidden limitations and the lack of transparency makes it difficult for tractography to be taken seriously. Studies trying to prove or disprove a hypothesis should clearly define their goal before launching processing. Questions such as *Must the conclusion be robust to “small” algorithmic choices?* need to be asked, otherwise positive results could be solely explained by a particularly severe bias, which could be eliminated by fine-tuning parameters (this point applies to pre-processing as well [Jones et Cercignani, 2010]).

When taking into account the full complexity of the dMRI tractography processing pipeline and the biases and pitfalls present in most, if not all, tractography algorithms, one could ask what is the usefulness of dMRI tractography? Tractography by itself cannot prove or disprove a hypothesis related to the underlying anatomy. However, it can provide useful insight that overall helps to understand the brain. It has opened a window into non-invasive structural brain mapping that neuroscientists, neurologists, radiologists and neurosurgeons have never had before. When the terminology, methodology, measurement error and analysis are clearly defined, clinical studies showing correlation with the intention of doing prediction or early detection are perfectly valid approaches.

## 2.5. ACKNOWLEDGEMENTS

However, any claim or phrasing trying to expose a causal link to the underlying anatomy or cognitive functions should be considered dangerous. The description of the limitations should have made that point clear by now. Once a hypothesis has been tested and reproduced using dMRI tractography, the only way to truly prove or disprove the said hypothesis is likely to be using an invasive (often ex-vivo) technique. However, considering the limited resources and the complexity of the human brain, dMRI tractography should be seen as a way to provide focus for invasive investigations.

Despite its own caveats, dMRI tractography has contributed to improving our understanding of the brain (and other organs). Future advances in diffusion acquisition, modelling and neuroanatomy will likely only increase the fundamental and clinical relevance of tractography. The goal of mapping the human brain “*Connectome*” is a multidisciplinary endeavor, and dMRI tractography is a needed tool that achieves it.

## 2.5 Acknowledgements

This work has been funded by the Fonds de recherche du Québec - Nature et technologies (FRQNT). A special thanks to Alexandre Gauvin for his work on tractogram and bundle saturation as well as Flavio Dell’Acqua for his work on the Wall effect. Thank you to the Neuroinformatics Chair of the Sherbrooke University which helped push forward neurosciences research.



“What is the point of being on this Earth  
if you are going to be like everyone else.”

---

– Arnold Schwarzenegger, Pumping Iron.

## Chapitre 3

# Tractostorm, le quoi, pourquoi et comment de la reproductibilité de la dissection en tractographie

### Résumé

Les recherches étudiant les structures de la matière blanche qui utilisent la tractographie par imagerie de résonance magnétique de diffusion requièrent fréquemment l'utilisation de segmentation manuelle de faisceaux, ou dissection virtuelle. Les erreurs humaines et les décisions personnelles font que ces segmentations manuelles sont difficiles à reproduire, opérations qui n'ont pas encore été quantifiées par la communauté de tractographie. Si la communauté de tractographie souhaite être prise au sérieux comme un outil reconnu en clinique, il est impératif d'harmoniser la segmentation de faisceau et de développer des protocoles destinés aux applications cliniques. Le *EADC-ADNI Harmonized Hippocampal Protocol (HarP)* a atteint ce niveau de standardisation à l'aide d'une séquence d'étapes rigoureuses qui devront être reproduites pour les différents faisceaux de la matière blanche. Cette publication est une observation de la problématique. Un protocole de segmentation visant un faisceau spécifique a été utilisé afin d'obtenir un exemple vraisemblable, mais la contribution de ce papier est de lancer la discussion concernant le besoin de protocole standar-

disé et reproductible. Ce projet a requis la participation de onze experts et treize non-experts en neuroanatomie et en dissection virtuelle provenant de laboratoires et hôpitaux un peu partout dans le monde. Plusieurs mesures liées aux voxels et aux *streamlines* ont été utilisées pour évaluer et comparer les résultats. Le protocole utilisé n'était pas nécessairement optimal, mais son design imite, en essence, ce qui sera requis d'un protocole futur. Des efforts coordonnés de la part de la communauté de diffusion sont nécessaires pour le design et la quantification de la reproductibilité de futurs protocoles de segmentation de faisceau.

## Contributions de la publication

- Formalisation d'un protocole de segmentation de faisceau destiné à n'importe quels participants.
- Investigation de métriques d'analyse liées à la comparaison de faisceaux.
- Agrégation des résultats de la première évaluation à grande échelle d'un protocole de segmentation de faisceau.
- Suggestion d'un cadre d'analyse rigoureux pour l'évaluation de futurs protocoles.

## Contributions des auteurs

- Conception de l'étude et recrutement des participants (Francois Rheault)
- Soumission du projet au comité d'éthique de l'Université de Sherbrooke (Francois Rheault)
- Travaux de segmentation manuelle (tous les auteurs sauf Francois Rheault et Maxime Descoteaux)
- Analyse, écriture et création et figures (Francois Rheault)
- Relecture et suggestions (tous les auteurs)
- Supervision du projet, aide à l'organisation et écriture (Maxime Descoteaux)

## Commentaires

À l'origine ce projet n'était qu'un moyen de quantifier les performances de l'algorithme RecobundlesX (voir Chapitre 4) par rapport à l'expertise humaine. Ce projet est rapidement devenu beaucoup plus complexe et intéressant que ce que j'avais anticipé. La soumission au Comité d'éthique de l'Université de Sherbrooke et le recrutement des participants ont été particulièrement difficiles. Malgré la nature de la publication, le journal Neuroimage (reconnu pour son accent méthodologique) a été initialement choisi pour son audience particulièrement large et diversifiée. Malheureusement, suite à des commentaires particulièrement négatifs qui ne voyaient pas l'intérêt du projet, la publication a été soumise à Human Brain Mapping. Ce journal, plutôt axé sur l'application et la recherche clinique possède aussi une audience très large, a mené à des commentaires très positifs sur l'importance de ce projet. (soumise 25 août 2019, en révision, 1er tour de réponse)

# Tractostorm : The what, why and how of tractography dissection reproducibility

Francois Rheault, MSc<sup>1</sup>, Alessandro De Benedictis, MD/PhD<sup>2</sup>, Alessandro Daducci, PhD<sup>3</sup>, Chiara Maffei, PhD<sup>4</sup>, Chantal M.W Tax, PhD<sup>5</sup>, David Romascano, MSc<sup>6</sup>, Eduardo Caverzasi, MD/PhD<sup>7</sup>, Felix C. Morency, MSc<sup>8</sup>, Francesco Corrivetti, MD<sup>9</sup>, Franco Pestilli, PhD<sup>10</sup>, Gabriel Girard, PhD<sup>6</sup>, Guillaume Theaud<sup>1</sup>, Ilyess Zemmoura, MD/PhD<sup>11</sup>, Janice Hau, PhD<sup>12</sup>, Kelly Glavin<sup>13</sup>, Kesshi M. Jordan, PhD<sup>7</sup>, Kristofer Pomiecko<sup>13</sup>, Maxime Chamberland, PhD<sup>5</sup>, Muhamed Barakovic, MSc<sup>6</sup>, Nil Goyette<sup>8</sup>, Philippe Poulin, MSc<sup>1</sup>, Quentin Chenot, MSc<sup>14</sup>, Sandip S. Panesar, MD/MSc<sup>15</sup>, Silvio Sarubbo, MD/PhD<sup>16</sup>, Laurent Petit, PhD<sup>17</sup>, Maxime Descoteaux, PhD<sup>1</sup>

<sup>1</sup>Sherbrooke Connectivity Imaging Laboratory (SCIL), Université de Sherbrooke, Sherbrooke, Canada,

<sup>2</sup>Neurosurgery Unit, Department of Neuroscience and Neurorehabilitation, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy,

<sup>3</sup>Computer Science Department, University of Verona, Verona, Italy,

<sup>4</sup>Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Boston, USA,

<sup>5</sup>Cardiff University Brain Research Imaging Centre (CUBRIC), School of Psychology, Cardiff University, Cardiff, United Kingdom,

<sup>6</sup>Signal Processing Lab (LTS5), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland,

<sup>7</sup>Department of Neurology, University of California, San Francisco, USA,

<sup>8</sup>Imeka Solutions, Sherbrooke, Canada,

<sup>9</sup>Département de neurochirurgie, Hôpital Lariboisière, Paris, France,

<sup>10</sup>Department of Psychological and Brain Sciences, Indiana University, Bloomington, USA,

<sup>11</sup>UMR 1253, iBrain, Université de Tours, Inserm, Tours, France,

<sup>12</sup>Brain Development Imaging Laboratories, Department of Psychology, San Diego State University, USA,

<sup>13</sup>Learning Research & Development Center (LRDC), University of Pittsburgh, Pittsburgh, USA,

<sup>14</sup>ISAE-SUPAERO, Toulouse, France,

<sup>15</sup>Department of Neurosurgery, Stanford University, Stanford, USA,

<sup>16</sup>Division of Neurosurgery, Emergency Department, "S. Chiara" Hospital, Azienda Provinciale per i Servizi Sanitari (APSS), Trento, Italy,

<sup>17</sup>Groupe d'Imagerie Neurofonctionnelle, Institut des Maladies Neurodégénératives - UMR 5293, CNRS, CEA University of Bordeaux, Bordeaux, France

**Keywords:** Diffusion MRI, White Matter, Tractography, Bundle segmentation, Intra-rater, inter-rater, Reproducibility m

## Abstract

Investigative studies of white matter (WM) brain structures using diffusion MRI (dMRI) tractography frequently require manual WM bundle segmentation, often called “*virtual dissection*”. Human errors and personal decisions make these manual segmentations hard to reproduce, which have not yet been quantified by the dMRI community. It is our opinion that if the field of

### 3.1. INTRODUCTION

dMRI tractography wants to be taken seriously as a widespread clinical tool, it is imperative to harmonize WM bundle segmentations and develop protocols aimed to be used in clinical settings. The EADC-ADNI Harmonized Hippocampal Protocol (HarP) achieved such standardization through a series of steps that must be reproduced for every WM bundle. This manuscript is an observation of the problematic. A specific bundle segmentation protocol was used in order to provide a real-life example, but the contribution of the paper is to discuss the need for reproducibility and standardized protocol, as for any measurement tool. This study required the participation of eleven (11) experts and thirteen (13) non-experts in neuroanatomy and “*virtual dissection*” across various laboratories and hospitals. Intra-rater agreement (Dice score) was approximately 0.77, while inter-rater was approximately 0.65. The protocol provided to participants was not necessarily optimal, but its design mimics, in essence, what will be required in future protocols. Reporting tractometry results such as average FA, volume or streamline count of a particular bundle without a sufficient reproducibility score could make the analysis and interpretations more difficult. Coordinated efforts by the diffusion MRI tractography community are needed to quantify and account for reproducibility of WM bundle extraction protocols in this era of open and collaborative science.

## 3.1 Introduction

DMRI tractography reconstructs streamlines that model the white matter (WM) neuroanatomy. The set of all streamlines forms an object often called the tractogram [Catani et De Schotten, 2008; Jeurissen et al., 2017]. When specific hypotheses about known pathways, i.e. WM bundles, are investigated, neuroanatomists design “*dissection plans*” that contain anatomical landmarks and instructions to isolate the bundle of interest from this whole brain tractogram [Bayrak et al., 2019; Catani et De Schotten, 2008; Catani et al., 2002; Chenot et al., 2019; Hau et al., 2016]. From now on “*dissection plans*” will be referred as segmentation protocols. Bundles can be segmented to study WM morphology, asymmetries, and then can be associated

### 3.1. INTRODUCTION

to specific functions [Catani et al., 2007; Groeschel et al., 2014; Lee Masson et al., 2017; Masson et al., 2018] with approaches similar to other brain structures [Lister et Barnes, 2009; Reitz et al., 2009]. Despite having similar anatomical definitions across publications, the absence of common segmentation protocols for tractography leads to differences that are for the most part unknown and unaccounted for. We need to know how variable our measurements are if we want to be able to have robust bundle-based statistics in the future. At the moment, there are no standardized method being used by the community.

The need for a gold standard that quantifies human variability is well-known and well-studied in other fields, such as automatic image segmentation, cell counting or in machine learning [Boccardi et al., 2011; Entis et al., 2012; Kleesiek et al., 2016; Piccinini et al., 2014]. For applications such as hippocampi or tumor segmentation, thorough assessments of reproducibility and multiple iterations of manual segmentation protocols already exist [Boccardi et al., 2015; Frisoni et al., 2015]. These protocols were specifically designed to reduce the impact of human variability and help outcome comparison in large-scale clinical trials across multiple centers [Frisoni et al., 2015; Gwet, 2012]. It is our opinion that the very same steps are needed for every WM pathways in order to achieve reproducible results. Our work is, in fact, an attempt to expose and clarify the necessity to design harmonized protocols, quantify their reproducibility and take variability into account when reporting results.

The reproducibility of manual bundle segmentation is likely to always be lower than manual image segmentation. Image segmentation in 3D requires local decision-making, and the decision to include voxels or not is directly done by raters. However, bundle segmentation requires local decisions that possibly impact the whole volume as streamlines reach outside of the scope of decisions made by raters. Since small or large hand-drawn regions of interest (ROI) or spheres are used to segment bundles, small mistakes can have far-reaching consequences. Even if ROIs are fairly reproducible in a strict protocol, the resulting bundles could be far from reproducible. This local-decision and global-impact conundrum makes the design of reproducible protocols more difficult and can potentially cause low agreement between raters.

## 3.1. INTRODUCTION

### 3.1.1 Bundle segmentation

Bundle segmentation is the action of isolating streamlines based on neuroanatomical priors, using known regions where certain conditions need to be satisfied. Inclusion and exclusion regions-of-interests (ROIs) are drawn and defined at the voxel-level using co-registered structural images, and are subsequently used to select the streamlines produced by tractography [Behrens et al., 2007; Catani et al., 2002; Ghaziri et al., 2015; Renauld et al., 2016; Rozanski et al., 2017], as seen in the Figure 3.1. Streamlines can be included or discarded using inclusion ROIs where streamlines are forced to traverse, and exclusion ROIs that cannot be crossed. Known structures such as grey nuclei, gyri or sulci and recognizable signal signatures can be used as landmarks to create a plan to follow for the segmentation [Catani et De Schotten, 2008; Catani et al., 2002; Chenot et al., 2019; Hau et al., 2016]. In this work, the person performing the task of segmentation (i.e drawing the ROIs, following the protocol) will be referred to as *rater*. Manual segmentation can be performed in software such as, but not limited to, DTI studio [Jiang et al., 2006], Trackvis [Wang et al., 2007], exploreDTI [Leemans et al., 2009], MITK Diffusion [Neher et al., 2012], FiberNavigator [Chamberland et al., 2014], or MI-Brain [Rheault et al., 2016] (Figure 3.1).

Once a bundle of interest is segmented from a tractogram, the analysis varies according to the research question. It is common to report asymmetry or group difference in bundle volume [Catani et al., 2007; Chenot et al., 2019; Song et al., 2014], diffusion values within the bundle of interest (average fractional anisotropy, mean diffusivity, etc.) [De Erausquin et Alba-Ferrara, 2013; Kimura-Ohba et al., 2016; Ling et al., 2012; Mole et al., 2016] or values along the bundle (called *profilometry* and *tractometry*) [Cousineau et al., 2017; Dayan et al., 2016; Yeatman et al., 2012, 2018]. Spatial distribution of cortical terminations of streamlines can help to identify cortical regions with underlying WM connections affected by a condition [Behrens et al., 2003; Donahue et al., 2016; Johansen-Berg et al., 2004; Mars et al., 2011; Rushworth et al., 2005]. Reporting the number of streamlines (e.g in connectivity matrix or density maps) is still very much present as a way to compare groups [Girard et al., 2014; Jones et al., 2013; Sotiropoulos et Zalesky, 2019], despite not being directly related to anatomy or connection strength [Jones, 2010; Jones et al., 2013].



### 3.1. INTRODUCTION

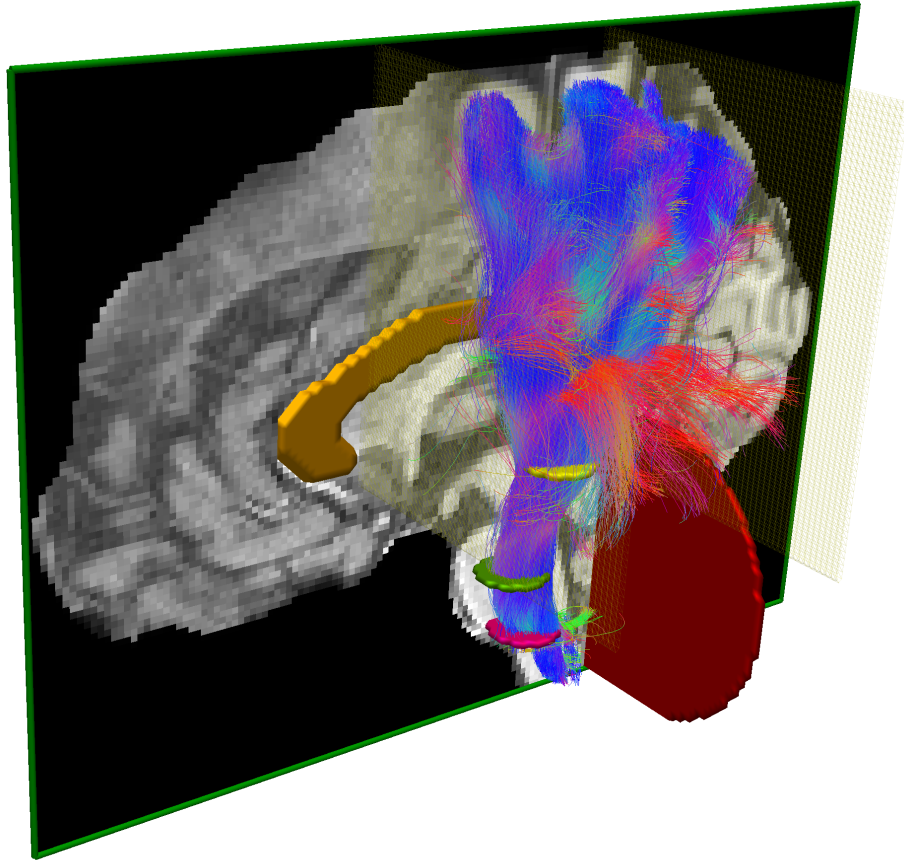


figure 3.1 – Illustration of the dissection plan of the PyT using the MI-Brain software [Rheault et al., 2016]. 3 axial inclusion ROIs (pink, green, yellow), 1 sagittal exclusion ROI (orange), 2 coronal exclusion ROIs (light yellow) and a cerebellum exclusion ROI (red, optional). The whole brain tractogram was segmented to obtain the left pyramidal tract.

#### 3.1.2 Quantifying reproducibility in tractography

When performing segmentation, it is crucial that raters perform the tasks as closely as possible to the dissection plan. Even if a single individual performs all segmentations, the possibility of mistakes or erroneous decisions about landmarks exists [Boccardi et al., 2011; Entis et al., 2012; Frisoni et al., 2015]. High reproducibility

### 3.1. INTRODUCTION

is often an assumption, if this assumption is false, the consequence could lead to inconsistent outcomes and erroneous conclusions. To assess the level of reproducibility of raters, identical datasets need to be segmented blindly more than once [Frisoni et al., 2015; Gisev et al., 2013; Gwet, 2012]. The literature on the subject, specifically for tractography, is quite sparse. Reported values for intra-rater and inter-rater variability are sometimes hidden in the methods section. However, it is common to report measures such as volume or average FA, which do not directly relate to spatial agreement (as detailed in Section 3.2.4), or to report variability of the ROI drawn by raters instead of the resulting bundles. Finally, the steps to perform the segmentation (the protocol) is not provided and the framework for evaluation is not defined. Despite these limitations, the general trend is that different bundles do not have the same level of variability (ranging from 0.4 to 0.95) and that algorithmic choices (ex. diffusion tensor imaging versus high angular resolution diffusion imaging) have an influence on variability [Colon-Perez et al., 2016; Dayan et al., 2015; Kaur et al., 2014; Kreilkamp et al., 2019; Voineskos et al., 2009; Wakana et al., 2007; Yendiki et al., 2011]. The first, and probably most complete, publication on the subject of protocol reproducibility assessment was from Wakana et al. [2007]. The acquisition and algorithmic choices for tractography could be considered sub-optimal (low spatial/angular resolution, diffusion tensor) and the framework inadequate for the now more common large-scale collaborations. The publication, despite providing a robust design, had a limited number of raters and duplicated data. The importance of reproducibility assessment, sparse literature on the subject and limited availability of protocols support the need for the work presented in this study.

To come back to tractography, the main message of our work is simple : Any study involving a manual segmentation protocol must provide a quantification of its measurement error, if it was never assessed before. Any modifications to the experimental setup will require a new assessment of the measurement error. Reporting measurements such as average FA, volume or streamline count, without a sufficient reproducibility is potentially problematic. Too low of an agreement score (e.g below 30%) could even mean an entirely different BOI/ROI is segmented, which would hinder further analysis. Similarly to the The EADC-ADNI Harmonized Hippocampal Protocol (HarP) [Frisoni et al., 2015], future dMRI tractography protocols will have

### 3.1. INTRODUCTION

to be designed for each bundle of interest. Groups of experts will have to propose protocols, pick-and-choose the best features of each and design an agreed upon set of rules, as it was undertaken for the HarP project [Boccardi et al., 2015]. As of this moment, efforts are being made toward an inter-protocol variability study to evaluate the current state of anatomical definitions present in the field, similar to the HarP project [Boccardi et al., 2011]. In this work, measurement error is sometimes referred as variability, while reproducibility is the capacity to reach the same results twice. Reproducibility of segmentations from the same individual is referred to as intra-rater agreement, while reproducibility of segmentation across raters is referred to as inter-rater agreement.

In the field of neuroimaging, voxels are used as the typical representation of data, while the available representation in tractography is in the form of streamlines (i.e. sets of 3D points in space). Figure 3.2 is a sketch of both representations. Several similarity measures exist to compare voxel-wise segmentations, e.g Dice score. Most of them have an equivalent formulation to compare sets of streamlines. However, resulting values can widely vary as the spatial distribution is not the same for both representations. Some measures related to streamlines require the datasets to be exactly the same, e.g Dice score, as streamline reconstructions are sets of discrete points with floating point coordinates and not discrete grids like 3D images. For this reason, comparison of streamlines is more challenging and datasets that do not originate from the same source distance in millimeters is often the only available solution [Garyfallidis et al., 2018; Maier-Hein et al., 2017].

### 3.1. INTRODUCTION

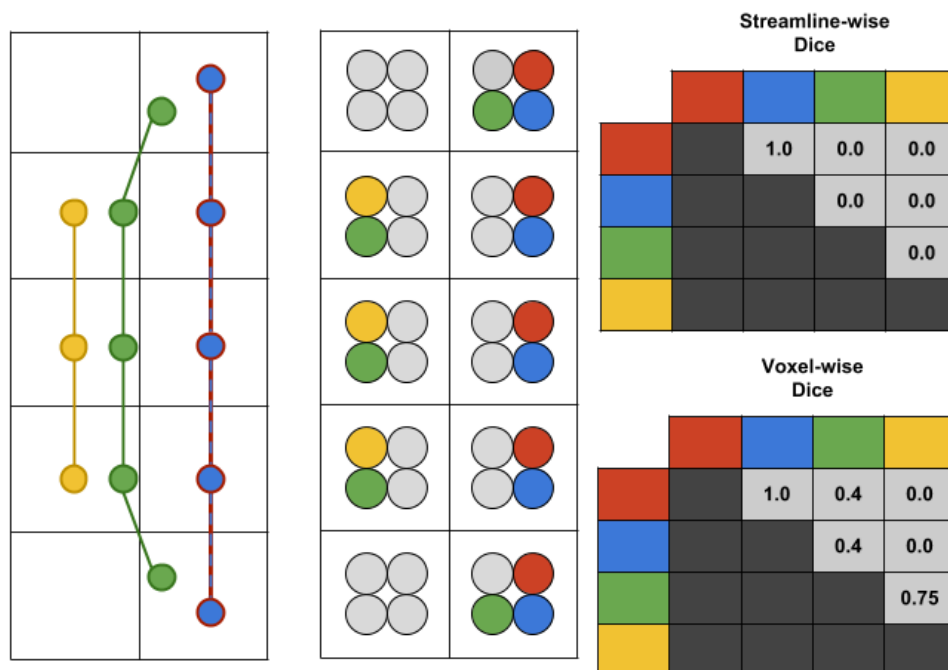


figure 3.2 – Representation of the Dice Coefficient (overlap) for both the streamline and the voxel representation. For the purpose of a didactic illustration, 4 streamlines are showed in a 2x5 “voxel grid”, the red and blue streamlines are identical. Each streamline is converted to a binary mask (point-based for simplicity) shown in a compact representation. Voxels with points from 3 different streamlines will results in voxels with 3 different colors, this can be seen as a spatial smoothing. The matrices on the right show values for all pairs (symmetrical). The green and yellow streamline are not identical, which results in a streamline-wise Dice coefficient of zero. However, in the voxel representation they have 3 voxels in common and the result is  $(\frac{2*3}{5+3} = 0.75)$ .

Automatic segmentation methods are becoming more widespread. Methods such as, but not limited to, [Chekir et al., 2014; Garyfallidis et al., 2018; Guevara et al., 2011; O’donnell et al., 2013; O’Donnell et al., 2017; Wassermann et al., 2016; Wasserthal et al., 2018a; Yendiki et al., 2011; Zhang et al., 2018a] aim to simplify the work of raters. The typical standard of most automatic segmentation method is to reach the accuracy of raters, thus it is crucial to truly quantify human reproducibility in manual tasks. It is possible to envision a scenario where an automatic method would not be as accurate as human expertise, but still useful to provide insight or even valid

### 3.1. INTRODUCTION

biomarkers. In such a case, it is still useful to know how accurate human expertise is, at least to provide comparisons.

#### 3.1.3 Summary of contributions of this work

Our PyT evaluation experiment serves as an “*example*” to convey the point that tractography segmentation is not standardized and needs to be addressed properly to make it quantitative, robust and more useful in the neuroimaging & human brain mapping literature.

We propose a framework to evaluate and quantify human reproducibility of bundle segmentation from dMRI tractography. Obtaining a measurement of rater (intra and inter) agreement is extremely relevant to set an appropriate threshold for statistical significance. It is also relevant for meta-analysis aiming to study large sets of publications and synthesize their outcomes. An account of human errors or other sources of variability is necessary. The main goal of this publication is to promote the importance of the assessment of any new “*virtual dissection*” protocol. We do not want to promote a given dissection protocol but emphasize the fact that any new protocol, including a new tractography algorithm and another bundle of interest would require a new assessment.

## 3.2. METHOD

A second contribution of this work is to investigate overlap, similarity measures and gold standard comparison designed for tractography. Development of easily interpretable measures for bundle comparison is necessary for large datasets. Overall the voxel representation is significantly more reproducible than the streamline representation. The voxel representation is better suited for analysis of tractography datasets (e.g reporting volume instead of streamline count). More details about these different representations and voxel/streamline-wise measures will be detailed in the [Method](#) and [Results](#) Section.

## 3.2 Method

### 3.2.1 Study design

Twenty-four participants were recruited and divided into two groups : experts and non-experts. The division was based on their neuroanatomical educational background. Participants working as researchers or PhD students in neuroanatomy, neurology or with extended experience in the field performing “*virtual dissection*” as well as neurosurgeons were part of the experts group (11 participants). The non-experts group was composed of MSc, PhD student or Post-Doc in neuroimaging, but without any formal education in neuroanatomy (13 participants). All participants had knowledge of dMRI tractography in general as well as the concept of manual segmentations of tractography datasets. Participation was voluntary and anonymous, recruitment was done individually and participants from various labs in Europe and the USA were solicited. The study was performed according to the guidelines of the Internal Review Board of the Centre Hospitalier Universitaire de Sherbrooke (CHUS).

Five independent tractograms and their associated structural/diffusion images were used, each was triplicated (total of 15). One was untouched, one was flipped in the X-axis (left/right) and one was translated. This was done to ensure that the participants were not aware they were performing reproducibility tasks. The symmetry of the segmentation plan (no difference between hemispheres) and lack of absolute frame of reference (coordinates) allowed these operations. Then, all datasets were randomly named so the tasks could be performed blindly for each participant. Parti-

### 3.2. METHOD

cipants were not aware of the presence of duplicated datasets. Five tractograms were used to obtain stable averages, duplicated datasets were used to score the intra-rater agreement and the multiple participants to evaluate inter-rater agreement. The decision to separate participants in two groups was made to generate additional data about reproducibility in real-life conditions.

Figure 3.3 shows an overview of the study design. To evaluate intra-rater reproducibility of rater #1, each triplicate was used to compute reproducibility measures. Meaning that 5 (A-B-C-D-E) x 3 (1-2-3) values were averaged to obtain the intra-rater “*reproducibility score*” of a single rater. To evaluate inter-rater reproducibility of rater #1, triplicates were fused and compared to all other raters to obtain a reproducibility measure. Meaning that 5 (A-B-C-D-E) x N (raters) values were averaged to obtain a single inter-rater “*reproducibility score*”. To evaluate reproducibility of rater #1 against the gold standard, the fused triplicates were also used. Meaning that 5 (A-B-C-D-E) x 1 (gold standard) values were averaged to obtain a single rater gold standard “*reproducibility score*”. The results shown in the Results Section are average values from all raters in each group. All reproducibility measures were computed using the same approach.

#### 3.2.2 DWI datasets, processing and tractography

Tractograms were generated from the preprocessed HCP [Van Essen et al., 2013] DWI data (3 males and 2 females, healthy, 26-35 years old) using three shells (1000, 2000, 3000) with 270 directions. The  $b_0$ , fractional anisotropy (FA) and RGB (colored FA) images were computed from DWI to be used as anatomical reference during segmentation. Constrained spherical deconvolution (CSD) using a FA threshold from a tensor fit on the  $b=1000s/mm^2$  was used to obtain fiber orientation distribution functions (fODF) [Descoteaux et al., 2007; Tournier et al., 2007] (spherical harmonic order 8) from the  $b=2000s/mm^2$  and  $b=3000s/mm^2$  shells. Probabilistic particle filtering tractography [Girard et al., 2014] was subsequently computed at 30 seeds per voxel in the WM mask (FSL FAST [Woolrich et al., 2009]) to make sure sufficient density and spatial coverage were achieved. The decision to use high-quality data was to maximize the quality of the tractogram so it would not be a limiting factor for the

### 3.2. METHOD

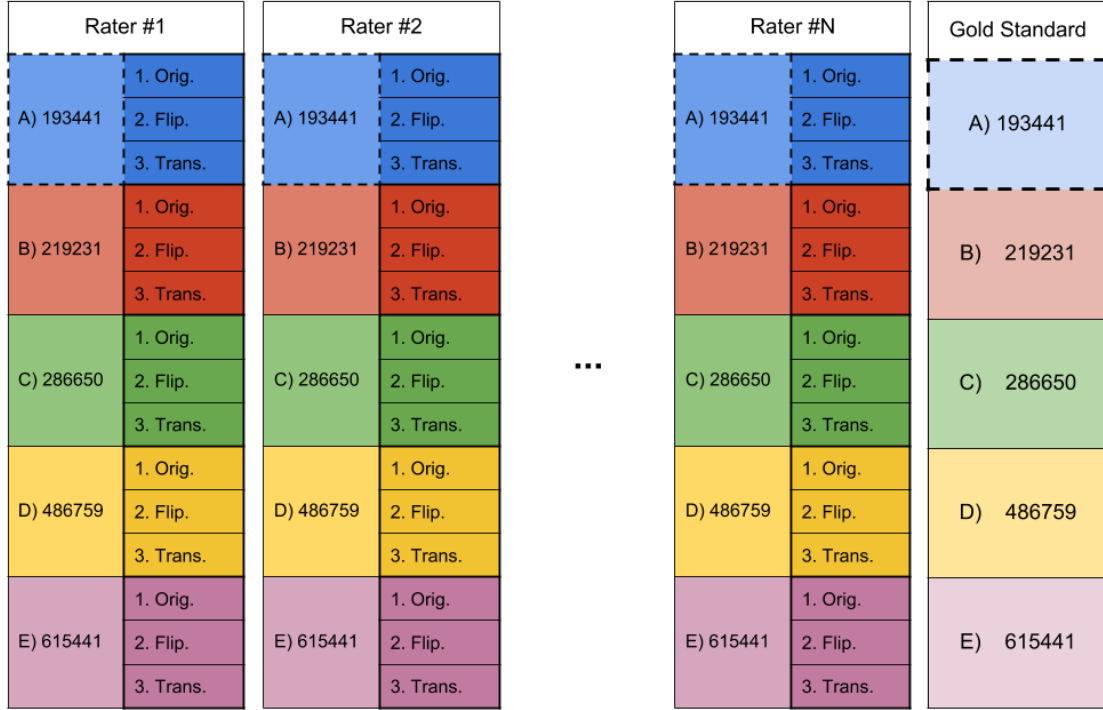


figure 3.3 – Representation of the study design showing N participants, each received 5 HCP datasets (listed and color-coded) which were replicated 3 times (original, flipped, translated). All participants had to perform the same dissection tasks, on the same anonymized datasets. Intra-rater, inter-rater and gold standard reproducibility were computed using the deanonymized datasets. More details are available in the supplementary materials

segmentation tasks. Moreover, it is now more and more common in clinical research to reach resolution as high as 1.5mm isotropic with multi-shell schemes, such dataset generates tractograms on a similar quality range as the HCP.

The CSD model was also used for bundle-specific tractography (BST) to further improve density and spatial coverage of the bundle of interest [Rheault et al., 2019]. This was to ensure that the full extent of the PyT was reconstructed and to ensure not to have criticisms from our experts in neuroanatomy complaining of a lack of fanning [Pujol et al., 2015]. A large model that approximates the corticospinal tracts (CST), which encompass the PyT, was used to generate streamlines with a strong preference



### 3.2. METHOD

for the Z-axis (up-down). A similar PyT reconstruction could have been achieved by generating millions and millions of streamlines, which would have been heavy and cumbersome from dissectionnists participants in the study (e.g computer performance requirements). This approach was used to increase the PyT reconstruction quality in [Chenot et al., 2019] and demonstrated its usefulness.

Furthermore, only the general orientational priors was used (globally helping the ‘*up/down*’ orientation) during tractography and not spatial/tissue priors was used. Meaning that globally results in a whole brain tractogram that was seeded from tens of thousands of voxels from a large region apparent to the CST. And these results were fused to a conventional whole brain CSD probabilistic tractogram. The resulting tractogram provided to the participants is indistinguishable from a conventional one. The rationale for this decision to use a more efficient seeding/tracking method to fill-up the full spatial extent of the PyT, as opposed to generating 10-50 millions of streamlines.

To accommodate all participants and the wide range of computer performance, tractograms were compressed using a 0.2mm tolerance error [Presseau et al., 2015; Rheault et al., 2017] and commissural streamlines were removed and datasets split into hemispheres. Each hemisphere (of each subject) had approximately 500 000 streamlines.

## 3.2. METHOD

### 3.2.3 Dissection plan and instructions

Each participant received their randomly named datasets, a document containing instructions for the segmentation and a general overview of a segmentation as an example (see supplementary materials). The segmentation task consisted in 15 segmentations of the pyramidal tract (left and right). The rationale behind the decision to focus on this PyT bundle was first, that a well-defined and well-known pathway was desired. Second, a dissection plan made of small and large inclusion and exclusion regions was desired. Finally, that the general shape was intuitive so the participants with no background in neuroanatomy could perform the tasks. Segmentation involved using 3 WM inclusion ROIs (Internal capsule, Midbrain and Medulla Oblongata) and 2 exclusion ROIs (one plane anterior to the precentral gyrus and one plane posterior to the postcentral gyrus). The detailed segmentation plan is available in the supplementary materials [Chenot et al., 2019] (<https://doi.org/10.5281/zenodo.2547024>).

Participants had to perform the segmentation plans, following the instructions as closely as possible. The dataset order was provided in a spreadsheet file. Participants had to choose between two software; Trackvis [Wang et al., 2007] or MI-Brain [Rheault et al., 2016]. This decision was made to guarantee that the data received from all participants was compatible with the analysis. Metadata such as date, starting time and duration had to be noted in the spreadsheet file. Upon completion, the participants had to send back the same 15 folders with two tractography files in each, the left and right pyramidal tract (PyT).

### 3.2.4 Bundles analysis

Once returned by all participants, datasets were de-randomized to match triplicates across participants. The duplicates (flipped and translated) were reverted back to their native space and all datasets (images and tractograms) were warped to a common space (MNI152c 2009 nonlinear symmetrical) using the Ants registration library [Avants et al., 2008; Fonov et al., 2011] to simplify the analysis. With all datasets having a uniform naming convention and in a common space, the intra-rater and inter-rater reproducibility can be assessed.

## 3.2. METHOD

### Individual measures

Reproducibility can be assessed using various measures. Average FA, volume and streamline count are the main attributes obtained directly from files. They do not provide direct insight about reproducibility, but one could expect that very similar segmentations should have very similar values. However, segmentation could wildly differ across rater and yet these measurements could be very similar. Average FA, volume or streamline count comparison do not provide any information about reproducibility. Reporting values from bundles obtained via manual segmentation using a protocol with unknown reproducibility scores is uninterpretable. This is simply due to the fact that completely different bundles can have the same measurements. This is why a confirmation that raters following the same protocol obtain the “*same*” segmentation is crucial, i.e high reproducibility. In this work results for the left and right PyT are averaged together without distinction, they are considered the same bundle during the analysis.

### Intra-rater and inter-rater

Each participant performed the same tasks on each triplicate. The goal of this triplication is to evaluate intra-rater reproducibility. Since all participants had access to the same datasets, inter-rater reproducibility can be assessed too.

Computing the average value from all pairwise combinations provides an estimate of the agreement between multiple segmentations of a same bundle. The deviation can also provide insights about the consistency of these segmentations. Measurement values can be between 0 and 1, such as Dice and Jaccard [Dice, 1945], meaning they are independent of the size. Figure 3.5 shows bundles and how to interpret these measures. Pearson’s correlation coefficient obtained from density maps provides insight into the statistical relationship and spatial agreement between two segmentations [Hyde et Jesmanowicz, 2012]. Each measure provides a way to interpret the data at hand, but there is no single true measure to summarize intra-rater and inter-rater agreement.

### 3.2. METHOD

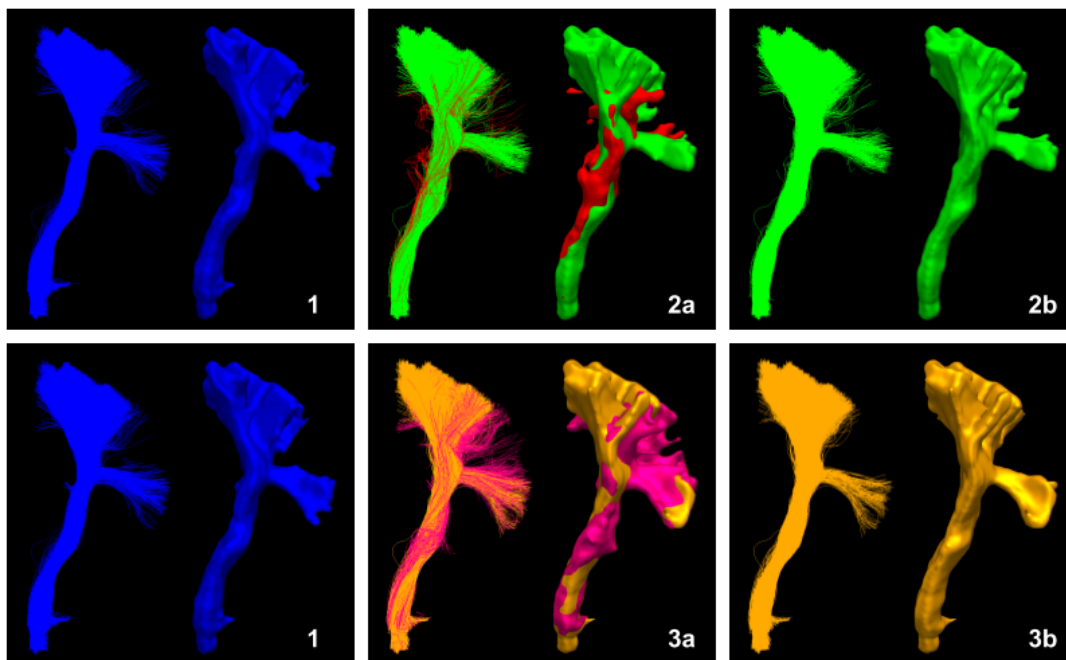


figure 3.4 – Comparison of bundles and the impacts of spurious streamlines on the reproducibility measurements. Each block shows streamlines on the left and the voxel representation on the right (isosurface). Block 2a and 3a shows the core (green/o-range) and spurious (red/pink) portion of the bundle. Block 2b and 3b only shows the core portion of the bundle.

		1-2a	1-2b	1-3a	1-3b
Correlation of the density map	VOX	0.90	0.91	0.93	0.94
Dice score	VOX	0.77	0.81	0.81	0.85
	STR	0.47	0.48	0.62	0.63

figure 3.5 – Table showing the reproducibility “score” between bundles, VOX marks voxel-wise measures and STR marks streamline-wise measures.

#### Gold standard

When multiple raters provide segmentations from an identical dataset, it is of interest to produce a gold standard. For a voxel representation, a probability map can be constructed, where each voxel value represents the number of raters that counted the voxel as part of their segmentation [Frisoni et al., 2015; Iglesias et Sabuncu, 2015;

### 3.2. METHOD

Langerak et al., 2015; Pipitone et al., 2014]. This can be normalized and then thresholded to obtain a binary mask representing whether or not the voxel was segmented by enough rater. A threshold above 0.5 is often referred to as a majority vote. The same logic can be applied to streamlines, each streamline can be assigned a value based on the number of raters that considered it part of their segmentation.

This can be seen in Figure 3.6 where increasing the minimal vote threshold reduces the number of outliers and overall size. In this work, the gold standard *does not* represent the true anatomy and should not be interpreted as such. It simply represents the average segmentation obtained from a tractogram. All elements that are not in a gold standard are true negatives and all the ones present are true positives. By construction, the gold standard does not contain false positives or false negatives. Binary classification measures are available such as sensitivity or specificity. However, several other measures are available and each are a piece of the puzzle leading to a more accurate interpretation [Chang et al., 2009; Garyfallidis et al., 2018; Schilling et al., 2018b].

To produce our gold standard a majority vote approach was used from the segmentations of the experts group, as their knowledge of anatomy was needed to represent an average version of the bundle of interest. In the context of this work, the *gold standard* is actually the average segmentation from experts. For simplicity, the expression *gold standard* was used as it is the best approximation of what can/could be achieved by our group of experts. The vote was set at 6 out of 11 and each of the 5 datasets got its own left and right gold standard. Since the representation at hand is streamlines (which can be converted to voxels), a streamline-wise and a voxel-wise gold standard was created. A majority vote approach is not necessarily optimal, but in the context where experts could not collaborate beforehand or after, this approach is adequate to obtain an average representation of the segmentation.

### 3.3. RESULTS

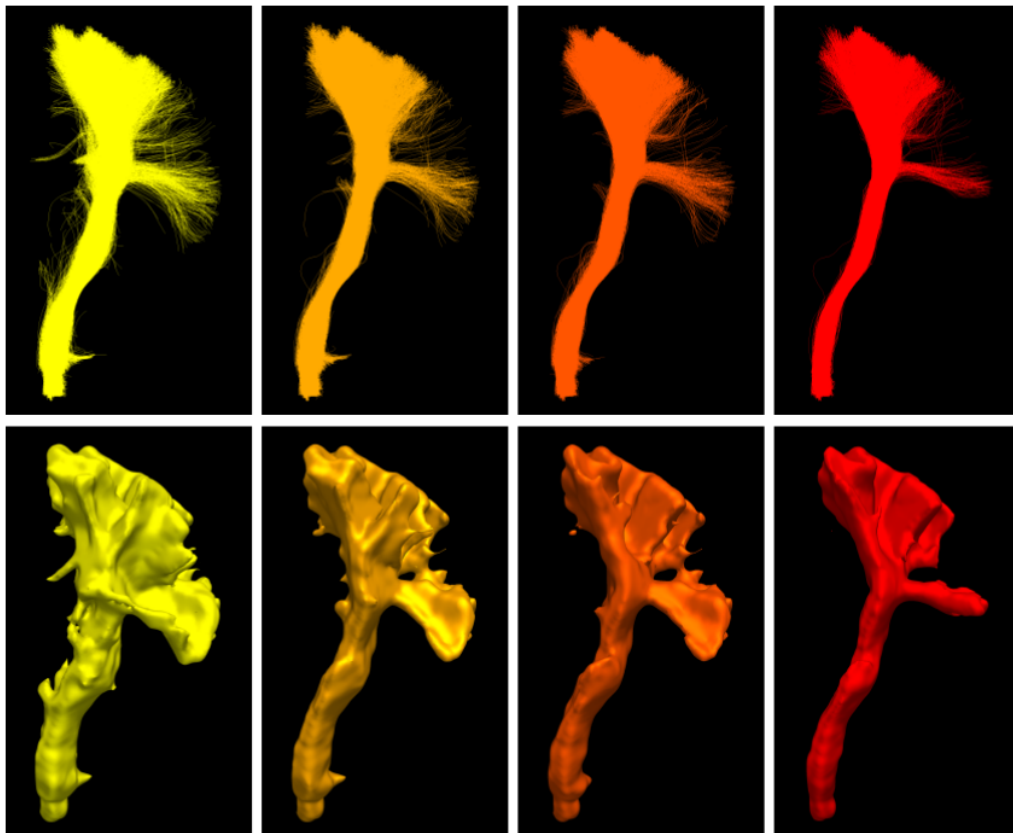


figure 3.6 – Example of average segmentation, or gold standard, generation obtained from 7 different segmentations, first row shows the streamline representation and the second row shows the voxel represented as a smooth isosurface. From left to right, multiple voting ratios were used ( $\frac{1}{7}, \frac{3}{7}, \frac{5}{7}, \frac{7}{7}$ ), each time reducing the number of streamlines and voxels consider part of the average segmentation. A minimal vote set at 1 out of 7 (left) is equivalent to a union of all segmentations while a vote set at 7 out of 7 (right) is equivalent to an intersection between all segmentations.

### 3.3 Results

In the following sections, all reported values are medians and interquartile ranges (IQR). This choice was made based on the fact that distributions are often bounded and not Gaussian distributions. Captions of figures report results as ( $Q_2$  (median);  $Q_3 - Q_1$  (IQR)), with a star (\*) indicating if the distributions are significantly different. All explicit comparisons between groups are statistically significant using a Mann-

### 3.3. RESULTS

Whitney rank non-parametric test for two independent samples ( $p - value < 0.01$ ).

On average, experts produce “*smaller*” bundles than non-experts, their volume and streamline count being lower than non-experts (-30% and -60%), as it can be observed Figure 3.7. This difference between groups is statistically significant ( $p - value < 0.01$ ). The range of values for segmentation measures is wider for non-experts, meaning that either intra-rater or inter-rater variability is likely higher. As mentioned earlier, this is useful insight about reproducibility, but lacks nuance and context. For example, despite obvious variation in volume and somewhat poor spatial overlap in segmentations (as shown in Figure 3.8 and 3.9), the average FA measurement does not show large variation. The fact that segmentations with low spatial overlap have the same average FA shows that reporting this measurement to gain insight about reproducibility of bundle segmentation is far from optimal. Since bundles without any overlap could have the same average FA, this measure is very difficult to interpret in terms of reproducibility evaluation.

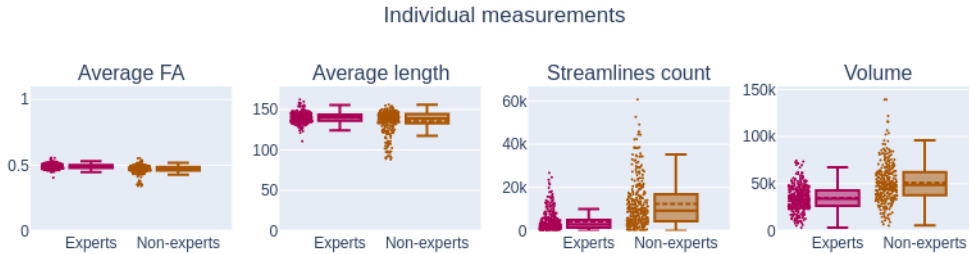


figure 3.7 – Measurements ( $Q_2$ ;  $IQR$ ) related to individual files for both groups. The Average FA distribution for experts (0.49; 0.01) and non-experts (0.47; 0.03) is not statistically different from each other. Similarly, the average length of experts (140.33mm; 7.81mm) and non-experts (138.70mm; 11.29mm) cannot be distinguished. Streamlines count of experts (2893; 3564\*) has a significant difference of distribution from non-experts (9383; 12368\*). The same can be same from the volume distribution (34.00cm<sup>3</sup>; 16.43cm<sup>3</sup>\*) for experts and (48.74cm<sup>3</sup>; 24.57cm<sup>3</sup>\*) for non-experts. The lower and higher fences for non-experts are much wider, indicating more variation in results.

### 3.3. RESULTS

#### 3.3.1 Intra-rater evaluation

All reported values can be seen in Figure 3.8. The median intra-rater overlap is represented by the voxel-wise Dice coefficient and is 0.75 for experts and 0.79 for non-experts. Streamline-wise Dice coefficient is much lower at 0.10 and 0.37 for both groups respectively. A higher Dice score value means that participants of a group are, on average, more reproducible with themselves. The median density correlation is equal ( $p - value < 0.01$ ) for both groups at 0.90 and 0.90 for the experts and non-experts group respectively.

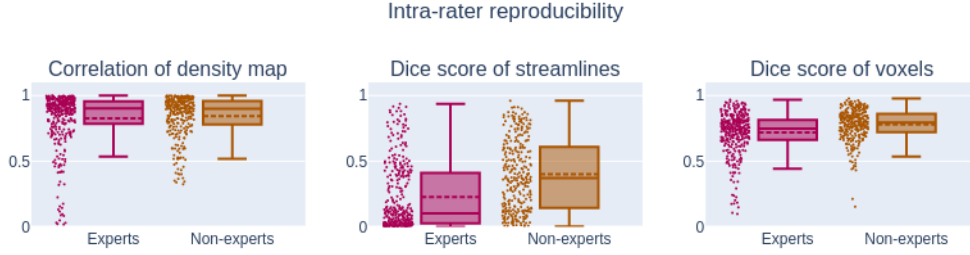


figure 3.8 – Measurements ( $Q_2; IQR$ ) related to pairwise comparison measures for intra-rater segmentations. The correlation of density maps showed no statistically significant difference between the experts (0.90; 0.17) and the non-experts (0.90; 0.17) groups. Distributions showed statistically significant difference for both Dice score. The Dice score of streamlines shows a easily observable difference between experts (0.10; 0.39\*) and non-experts (0.37; 0.46\*). The difference between distribution Dice score of voxels is less noticeable at (0.75; 0.15\*) for experts and (0.79; 0.14\*) for non-experts. The trend for the intra-rater reproducibility is that rater fail to select the same streamlines, but the ones that are selected still cover approximately the same volume.

#### 3.3.2 Inter-rater evaluation

To minimize the influence of intra-rater reproducibility during the evaluation of inter-rater reproducibility, the triplicate datasets were fused into a single bundle. This was performed to approximate the results as if participant segmentations had no intra-rater variability. This lead to an underestimation of inter-rater variability, but necessary to separate sources of variability later in the analysis. Voxel-wise Dice



### 3.3. RESULTS

coefficient is on average higher between experts than between non-experts, at 0.62 and 0.67 respectively while the streamline-wise Dice coefficient is much lower at 0.11 and 0.18. The median density correlation is higher between experts at 0.88 while non-experts are at 0.71. The IQR is higher for the non-experts group, meaning that the similarity among non-experts is not only lower, but widely varies. All reported values can be seen in Figure 3.9.

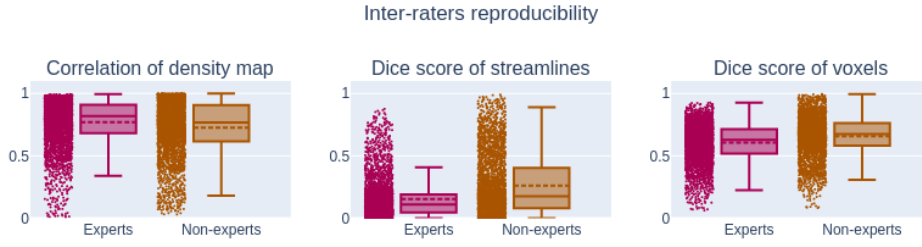


figure 3.9 – Measurements ( $Q_2$ ;  $IQR$ ) ( $Q_2$ ;  $IQR$ ) related to pairwise comparison measures for inter-rater segmentations. The correlation of density maps showed no statistically significant difference between the experts (0.82; 0.23\*) and the non-experts (0.77; 0.29\*) groups. Similarly to the intra-rater segmentation, distributions showed statistically significant difference for both Dice score. The Dice score of streamlines shows a easily observable difference between experts (0.11; 0.14\*) and non-experts (0.18; 0.32\*). While the distribution Dice score of voxels for experts (0.63; 0.20\*) and non-experts (0.67; 0.18\*) is more similar. Raters have difficulty to select the same streamlines, but overall capture similar volume.

#### 3.3.3 Gold standard evaluation

All reported values can be seen in Figure 3.10 and 3.11. Comparisons to the computed gold standard show that on average experts and non-experts obtain segmentation roughly similar to the average segmentation. However, all measures show that segmentations from experts are on average closer to the gold standard than those of non-experts. This was expected as the gold standard was produced using segmentations from the experts group. Values for streamline-wise measures are lower for Dice coefficient and density correlation, meaning that reproducibility is harder to achieve using the streamline representation. This was a similar trend observed in intra-rater and inter-rater values.

### 3.3. RESULTS

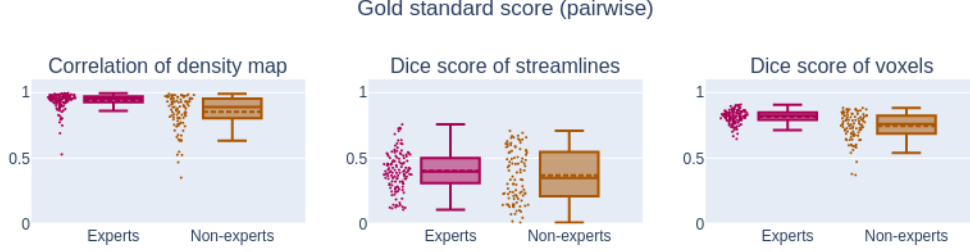


figure 3.10 – Measurements ( $Q_2$ ;  $IQR$ ) related to pairwise comparison measures against the gold standard. The correlation of density map reaching (0.95; 0.04\*) for experts and (0.88; 15\*) is statistically different between both groups. However, the Dice score of streamlines are not statistically different at (0.39; 0.18) and (0.34; 0.34) respectively. The Dice score of voxel is relatively high at (0.82; 0.05\*) for experts and (0.76; 0.13\*) for non-experts. Despite variations between rater, overall the participants remain around the same average segmentation and obtain more agreement with the gold standard than with each other.

Specificity and accuracy reach above the 95% for both groups both for streamlines or voxels. Meaning that experts and non-experts alike classified the vast majority of true negatives correctly. Since specificity is near a value of 1.0, the Youden score is almost equal to sensitivity. All 3 measures take into account the true negatives, which far outweigh the true positives, in our datasets, for this reason they were removed from Figure 3.11 and shown only in the supplementary materials. Sensitivity is much lower at 0.78 and 0.82 for experts and non-experts respectively, as both groups partially capture the gold standard. Precision is higher for experts (0.92) than for non-experts (0.78), meaning that experts were providing segmentations approximately the same *size* as the gold standard while non-experts were providing much bigger segmentations (that generally encompass the gold standard). This explains the higher sensitivity and lower specificity of non-experts. The Kappa and Dice score is lower for experts at 0.83 and 0.62 while the non-experts median is 0.79 and 0.67, respectively. Given the dimensionality of our data, getting the right segmentation by accident is very low, explaining why the Kappa and Dice score are very similar. It is important to consider that the ratio of true negatives to true positives is not the same for both representations (voxels vs. streamlines).

### 3.3. RESULTS

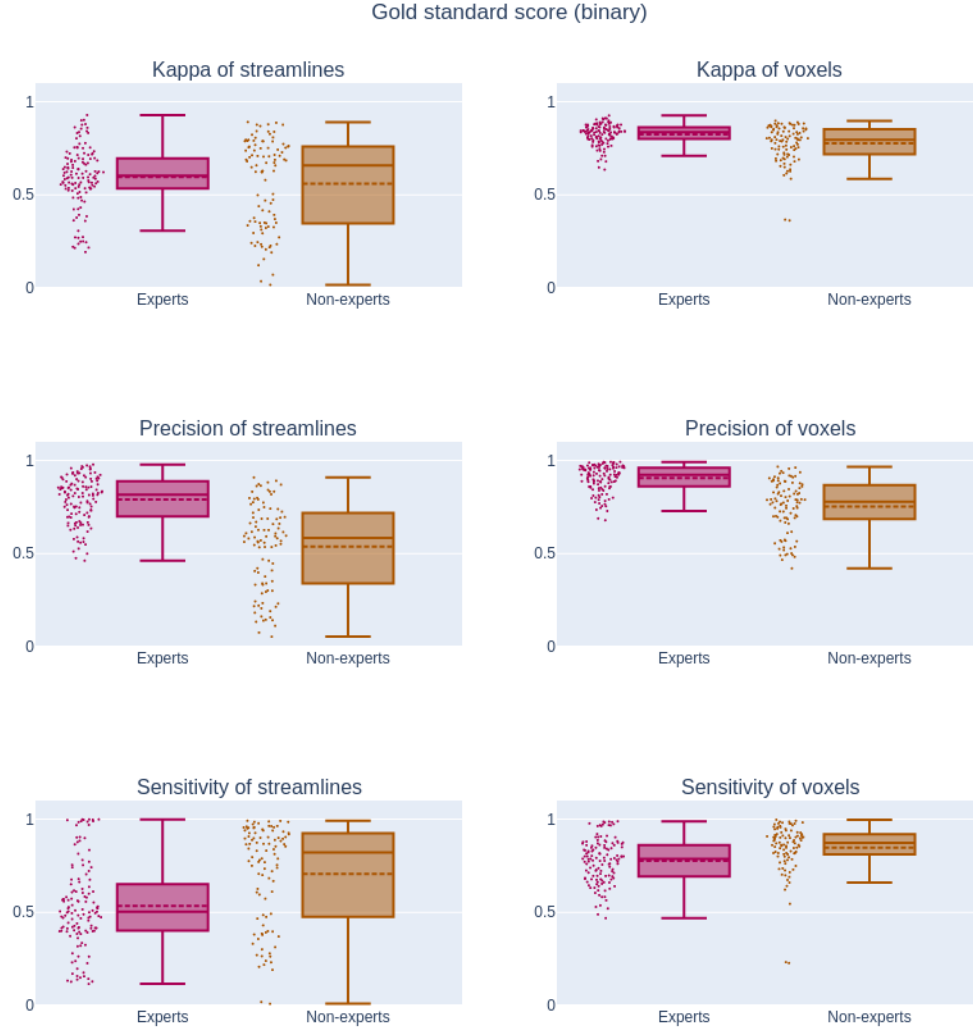


figure 3.11 – Measurements ( $Q_2$ ;  $IQR$ ) related to binary classification measures against the gold standard. The Kappa score is only significantly different for voxel (0.84; 0.06 and 0.80; 0.13) and not for streamlines (0.60; 0.16\* and 0.65; 0.41\*). There is a high degree of variability for precision and sensitivity of streamlines (0.81; 0.19\* and 0.50; 0.24\* for experts) and (0.59; 0.37\* and 0.82; 0.44\* for non-experts). These measures are more reliable with the voxel representation (0.92; 0.10\* and 0.79; 0.17\* for experts) and (0.78; 0.17\* and 0.82; 0.44\* for non-experts). The streamline representation is always less reproducible than the voxel representation. The measures such as accuracy and specificity are not shown due to the fact that both reach above 0.99 and do not provide useful visual insight.

### 3.4. DISCUSSION

The computation of inter-rater reproducibility was performed using the fused triplicate to minimize the influence of intra-rater reproducibility. The approach to fuse the triplicate is an approximation, fusing more than 3 segmentations from the same datasets would be necessary to perfectly evaluate intra-rater reproducibility. However, the 5 datasets used for this study represent sufficiently similar tasks to consider our approximation adequate for this work. Preliminary analysis showed low correlation values, between a participant “*score*” for intra-rater reproducibility and inter-rater reproducibility. Correlation was between 0.2 and 0.4 for all measures, this indicates that there is no clear link between the reproducibility of a participant’s own segmentations and the agreement with other participants.

## 3.4 Discussion

### 3.4.1 Evaluation of protocols

This work illustrates that intra-rater and inter-rater agreement is far from perfect even when following a strict and “*simple*” segmentation protocol. The intra-rater and inter-rater agreement must be taken into account when researchers compare bundles obtained from manual segmentations. When human expertise is required for a project, it is crucial that people involved in the manual segmentation process evaluate their own reproducibility, even if they have sufficient neuroanatomy knowledge and extensive experience in manual segmentation. This measure of error could increase the threshold for statistical significance. In such case, either more datasets will be needed, or a better protocol for segmentation needs to be designed [Boccardi et al., 2015; Gwet, 2012]. The similarity between both groups indicates that with the right protocol, it would be possible to train people without anatomical background to perform tasks with results and quality similar to experts.

Without such evaluation it is impossible for experts and non-experts alike to know beforehand how reproducible they are. Since their “*scores*” vary with the protocol, the bundle of interest and possibly other factors, it is important to consider an evaluation before performing large-scale segmentation procedure [Frisoni et al., 2015]. An alternative to guarantee more reproducible results is to design an appropriate protocol for

### 3.4. DISCUSSION

non-experts and to perform tasks blindly more than once. The results can then be averaged, which will make outliers and errors easier to be identified. Various ideas can be considered to facilitate the tasks and increase reproducibility. Using multiple modalities, such as functional MRI or myelin map, to identify important landmarks or allowing for wider ROIs delineation by adding a safe margin around each of them could help increase reproducibility. However, requiring additional modalities would make segmentation more specific, it may be of interested to segment a bundle based only structural data (T1, dMRI). However, any modifications to the current protocol, or any protocol for that matter, would create the need for a new reproducibility assessment using a similar framework to the one presented in this work. The conclusion remains the same, the lack of quantification about the reproducibility of a protocol is by itself problematic.

This study did not allow for collaboration and did not micro-manage participants, meaning they were left with the instructions without further intervention from the organizers. In a situation where a segmentation plan can be defined in groups and techniques can be improved along iterations of the plan, the intra-rater and inter-rater agreement would likely go up. This study aimed at the evaluation of participants following instructions from a protocol, similar to the ones present in books, publications or online examples.

#### 3.4.2 Measures and representations

In this work the intra-rater agreement was higher for non-experts than experts, without more information we could have concluded that non-experts were more meticulous when they were performing their manual segmentations. However, by looking at sensitivity and precision we can see that non-experts had “*bigger*” segmentations. Experts are likely stricter in their decision-making process, this could amplify the local-decision and global-impact conundrum mentioned earlier. A more liberal, less rigid, segmentation likely makes it easier to be reproducible, but does not necessarily make it valid. This is an example showing the importance of having more than one type of measure to obtain a complete picture.

### 3.4. DISCUSSION

In tractography, it is common to use a single measure to portray a complex phenomenon. Most measures used are simplified to have easily interpretable results. The previous example shows the importance of using more than one type of measurements to obtain a complete picture of the reproducibility. “*Reproducibility scores*” are likely to vary with the project and the bundle of interest. This needs to be addressed as a community. The discrepancy between protocol quality, reproducibility and conclusion put forward in the literature can be problematic.

For binary measures (accuracy and specificity), scores were both above 95% as it is easy to discard true negatives, and consequently did not provide much insight. Similarly to the curse of dimensionality in machine learning [Ceotto et al., 2011; Verleysen et François, 2005], our datasets typically contain millions of voxels (or streamlines), of which only a few thousands true positives are considered during segmentation. Thus, the vast majority of true negatives are rapidly discarded resulting in both accuracy and specificity almost reaching 100%. Sensitivity provides more information, as true positives are more difficult to get, since they are rarer in the tractograms (few thousands out of millions) [Maier-Hein et al., 2017]. This needs to be taken into account using precision, as in some cases, strict segmentation is encouraged because false positives are more problematic than false negatives. Streamline-wise measures show lower agreement, meaning that reproducible results are likely more difficult to achieve with the streamline representation.

More complex measures need to be designed to fully capture the complexity of tractography datasets and compare them, even across datasets or for longitudinal studies. Currently, more advanced measures that capture fanning, spatial coherence, localized curvature and torsion or spectral analysis are still rare, despite being used in other neuroimaging fields [Cheng et Basser, 2018; Esmaeil-Zadeh et al., 2010; Glozman et al., 2018; Lombaert et al., 2012].

### 3.4. DISCUSSION

#### 3.4.3 Tractography algorithms

Manual segmentation of deterministic tractograms is likely more reproducible, since small differences in ROI placement will have a smaller impact on the resulting bundle. The local-decision and global-impact conundrum mentioned earlier is more obvious with probabilistic tractography. Other tractography algorithms, such as global tractography [Christiaens et al., 2015; Kreher et al., 2008; Mangin et al., 2013; Neher et al., 2012], are likely to have different “*reproducibility scores*”, even with the same segmentation protocol. Any change to the preprocessing could lead to unexpected change in the reproducibility “*scores*”. For the purpose of this study, we provided optimized tractograms for the BOI and our algorithmic reconstruction choices may have influenced our results. However this is in line with our main message, which is that every project involving manual segmentation should come with its own reproducibility assessment. Hence, changing tractography algorithm would require a re-evaluation of the reproducibility is considered as part of important future work. For example, the project “*TractEm*” [Bayrak et al., 2019] featured a framework to obtain 61 bundle of interests from deterministic tractography and report some voxel-wise measures for intra-rater and inter-rater reproducibility. However, this protocol is likely optimized for specific datasets (BLSA, HCP). It also requires registration and tractograms must be generated with DSI-Studio (deterministic) only. Any deviation from this protocol would likely change the reported reproducibility measures.

Using the same dataset and tractography algorithm, but increasing or decreasing the number of streamlines, variation in step-size or angular threshold could also change the reproducibility “*scores*”. Another anatomical definition of the PyT, having that definition taught to participants in person instead of a simple PDF document or dissections of another BOI would likely lead to different reproducibility “*scores*”. Other dataset could come with their own challenges, e.g infants or aging population, where finding anatomical landmarks could be harder and thus lead to lower reproducibility.

Such trend can be observed in numerous other studies where investigation of the same bundle, different bundles or when different algorithmic choice leads to a wide variety of reported reproducibility “*scores*” [Colon-Perez et al., 2016; Dayan et al.,

### 3.4. DISCUSSION

2015; Kaur et al., 2014; Kreilkamp et al., 2019; Voineskos et al., 2009; Wakana et al., 2007; Wassermann et al., 2016; Yendiki et al., 2011]. Similar generalization difficulties and trends exist in the field of medical image segmentation [Boccardi et al., 2011; Frisoni et al., 2015]. However, the general conclusion remains that reproducibility needs to be quantified for each specific project and protocol. Reproducibility “scores” cannot be easily generalized and any attempt would be dangerous, as any deviation from a known protocol creates the need for a new assessment. Aiming for standardized and harmonized protocols that are agreed upon within the field should be the main focus on the long term, such as [Bayrak et al., 2019; Catani et De Schotten, 2008]

#### 3.4.4 Impact on analysis

If variability needs to be minimized further than the defined protocol, a simple recommendation is to have a single rater performs each task multiple times or multiple raters perform each task multiple times (or a subset of tasks). This way, it is guaranteed that each dataset is segmented more than once, decreasing the error risk. Regardless of the decision made, it is of great importance to quantify the reproducibility of manual segmentation of raters involved in the project before doing any statistics or group comparisons. This could drastically change the statistical significance of results. As of now, to the best of our knowledge, human variability and errors are rarely considered. Measurements such as volume and streamlines count can take into account the measurement error (i.e voxel-wise or streamline-wise Dice coefficient) as part of group variances. Combining the measurement errors with a group average can be achieved by using, for example, the principle of pooled variance [Gwet, 2012; Peters, 2001]. However, if the intra-rater “*reproducibility score*” is too low, e.g below 30%, reporting such measurements is counterproductive, as this will require much larger cohorts to reach statistical significance. Taking into account the measurement error makes sense as long as the same “*thing*” is being measured.

Sources of variability need to be accounted for to truly enable synthesis of work across multiple centers. Even when automatic or semi-automatic methods are used, they first need to be evaluated with agreed upon measures and reach or surpass human standards. In a very thorough longitudinal large-scale project across centers



### 3.4. DISCUSSION

involving manual segmentation, it would be desirable to acknowledge the variability across timepoints, across scanners and across rater into the analysis.

The extension to other bundles of interest or other segmentation plans is not trivial and the only conclusion that stands is that agreement is never 100% and that a unique measure is not enough to represent the whole picture for tractography segmentation. The desire to simplify measures or have only one value to describe quality or reproducibility of segmentations needs to be discouraged. The nature of our datasets makes this task much more complex to interpret than 2D or 3D images, and it is imperative that the field comes to understand and agree on measures to report. This is more relevant than ever as the field grows and now that open science is becoming more popular and reproducibility studies are encouraged. Similarly to other neuroimaging fields, such as hippocampi segmentation, standardized protocols need to be developed and designed to be used across multiple centers without active collaboration or micromanagement.

#### 3.4.5 Future work

Future work includes the creation of a database containing bundle segmentations and metadata from participants that will be available online so further analysis can be done. This metadata could help explain the variability, similar trend have been observed for most types of measurements, medical images or tractography segmentation [Boccardi et al., 2011; Bürgel et al., 2009; Gwet, 2012]. As for now, a preliminary upload of the participants segmentation is available on Zenodo (<https://doi.org/10.5281/zenodo.2547024>), which will be updated. In this work, metadata was not used to evaluate duration as a variable influencing reproducibility. Investigating the relationship between variability and duration of a task or looking for bias (inter-hemispheric or software influence) could be of interest for future research. An online platform similar to the Tractometer [Côté et al., 2013] or a Nextflow pipeline [Di Tommaso et al., 2017] is planned to be released. Such a tool would be designed for researchers to quickly submit data that is expected to have some level of agreement and obtain their “*reproducibility score*”. This way protocols can be improved and reproducibility can be taken into account in the analysis.

### 3.5. CONCLUSION

Protocols for many bundles need to be developed for various purposes, such as clinical practice, synthesis of findings, building training sets for machine learning, etc. The segmentation plan and instructions need to be defined clearly by panels of experts, and agreed upon terminology [Mandonnet et al., 2018], to optimize reproducibility and anatomical validity. The field of manual tractography segmentation is decades behind fields such as grey nuclei or hippocampi manual segmentation on this matter. The latter has been refining segmentation protocols for the past decade and has already reached the state harmonized segmentation protocol and was evaluated with reproducibility in various settings [Apostolova et al., 2015; Boccardi et al., 2015, 2011; Frisoni et al., 2015; Wisse et al., 2017].

## 3.5 Conclusion

When trying to understand how similar WM bundles from dMRI tractography are, at least 3 values need to be taken into consideration : *Dice coefficient of voxels* showing how well the overall volume overlaps, *Dice coefficient of streamlines* showing if the same streamlines were selected/discarded and *correlation of density map* showing if the streamlines are spatially distributed in a similar way. Results specific to our work on the pyramidal tract revealed that rater overlap is higher for voxel-wise measures (approximately 70%) than streamline-wise measures (approximately 20%).

In comparison to the group average, the results depict an ease to identify true negatives, an adequate number of true positives, while having a low number of false positives. The voxel and streamline representations do not produce equal levels of reproducibility. Studies reporting bundle asymmetry in terms of streamline count (streamline-based) will require a larger group difference than those reporting volume difference (voxel-based). Our particular protocol served as a powerful illustration of the importance of assessing the variability of human expertise when comparing population and provides interesting insights on WM manual segmentation.

The lack of framework for reproducibility assessment, the sparse literature on intra-rater and inter-rater variability in tractography and the variation in the reported

### 3.5. CONCLUSION

values across bundles, reconstructions, datasets and other variables points to the importance of our proposed framework for evaluation, as a step forward. It is of importance to re-iterate that the intention of this study is not to propose/enforce a processing pipeline for tractography and/or propose a new set of rules for PyT segmentation. The diversity of reported values for reproducibility shows that clearly defined processing and segmentation protocol is necessary in this field. All data and metadata used in this work is now publicly available (<https://doi.org/10.5281/zenodo.2547024>) in the hope to stimulate discussions and more evaluations in the future for other bundles and protocols. Better reproducibility of results is needed and goes hand-in-hand with the open science movement. A collaborative effort to evaluate and quantify human variability is needed.

## Acknowledgements

A special thanks to the funding sources for this work, the Fonds de recherche en santé Québec - Nature et technologies (FRQNT) and Collaborative Research and Training Experience Program in Medical Image Analysis (CREATE-MIA) programs. Thank you to the Neuroinformatics Chair of the Sherbrooke University which helped push forward neuroscience research.

## Data Availability Statement

The data that support the findings of this study are openly available in on Zenodo at <https://doi.org/10.5281/zenodo.2547024>.

“Don’t start chasing applause and  
acclaim. That way lies madness.”

---

— Ron Swanson, Parks&Recreation.

# Chapitre 4

## Segmentation automatique de faisceaux par approches multi-atlas avec fusion d'étiquettes

### 4.1 Introduction

Lorsqu'un tractogramme complet est reconstruit, l'objectif est de le décomposer en faisceaux pour faciliter l'analyse. L'une des approches les plus communes est d'utiliser des *a priori* anatomiques pour segmenter le tractogramme en faisceaux d'intérêt. Cette opération, nommée dissection virtuelle, est encore à ce jour effectuée manuellement [Bayrak et al., 2019; Catani et De Schotten, 2008; Catani et al., 2002; Chenot et al., 2019; Hau et al., 2016]. L'objectif de ces segmentations est d'isoler un faisceau et d'étudier sa morphologie, des possibles asymétries et ainsi de lier celui-ci à des fonctions spécifiques [Catani et al., 2007; Groeschel et al., 2014; Lee Masson et al., 2017; Masson et al., 2018]. De nombreuses méthodes automatiques ont depuis fait leur apparition dans le but de faciliter la tâche des neuroanatomistes, mais aussi dans l'espoir de réduire la variabilité expérimentale liée aux manipulations effectuées par des humains [Chekir et al., 2014; Garyfallidis et al., 2018; Guevara et al., 2011; O'donnell et al., 2013; O'Donnell et al., 2017; Wassermann et al., 2016; Wasserthal et al., 2018a; Yendiki et al., 2011; Zhang et al., 2018a].

## 4.1. INTRODUCTION

La base de l’algorithme présenté dans ce chapitre provient d’un projet antérieur du laboratoire proposé par [Garyfallidis et al. \[2018\]](#), mais ce projet (*Recobundles*) est trop brut pour être utilisé à grande échelle de façon satisfaisante. En effet, la méthode est à risque de produire des résultats très bruités et est sensible à la modification d’hyperparamètres. L’objectif de ce chapitre est de présenter une extension multi-atlas et multi-paramètres avec fusion d’étiquettes (*RecobundlesX*) inspirée d’algorithmes de segmentation d’imagerie médicale.

La méthode *Recobundles* repose sur une association par similarité entre *streamlines* provenant du tractogramme et d’un modèle. Un point central de la méthode *Recobundles* est la métrique de distance entre les *streamlines*, la *minimum direct flip distance* (MDF), qui est utilisée pour sa simplicité computationnelle [[Garyfallidis et al., 2010](#)]. Cette mesure de distance permet ensuite de grouper les *streamlines* par similitude à l’aide de l’algorithme *Quickbundles* [[Garyfallidis et al., 2012](#)]. Grâce à cet algorithme de partitionnement ou *clustering*, une méthode permettant le recalage entre faisceaux, la *streamlines-based linear registration* (SLR), a été développée [[Garyfallidis et al., 2016](#)]. Ces innovations subséquentes, élaborées sur plusieurs années, ont mené au développement de *Recobundles*, cette méthode est généralement robuste et suffisamment rapide pour avoir un intérêt pour des applications cliniques, mais plusieurs lacunes sont toujours présentes. La publication initiale s’est concentrée sur la présentation de ces étapes pour ensuite démontrer l’utilité de la méthode grâce à des expérimentations variées, mais limitées par une analyse en surface.

La principale contribution de ce chapitre est la création d’une approche multi-atlas, l’exploration d’un ensemble de paramètres d’entrée, et intégrant une fusion d’étiquettes par vote majoritaire. Une contribution secondaire est une analyse en profondeur des performances de *Recobundles* et *RecobundlesX* à l’aide d’un grand ensemble de données issues du projet *Tractostorm* (voir Chapitre 3 ou [Rheault et al. \[2020\]](#)). Des contributions indirectes issues de l’utilisation de *RecobundlesX* dans le cadre de projets sur le vieillissement, la sclérose en plaques, les commotions cérébrales, les accidents vasculaires cérébraux et la neurochirurgie sont aussi à prendre en compte.

## 4.1. INTRODUCTION

### 4.1.1 Segmentation multi-atlas de faisceaux

L'état de l'art en segmentation de structures cérébrales sur des images anatomiques repose une technique de fusion d'étiquettes provenant d'une multitude d'atlas ayant été segmentés manuellement [Collins et Pruessner, 2010; Iglesias et Sabuncu, 2015; Lampert et al., 2016; Pipitone et al., 2014]. Nous voulons utiliser une technique similaire pour la segmentation des principaux faisceaux de la matière blanche. L'extension de *Recobundles* à une version supportant plusieurs paramètres et plusieurs atlas permettra d'améliorer et de stabiliser les résultats, de façon similaire aux approches en segmentation d'image.

L'amélioration de la méthode *Recobundles* passe par le support de plusieurs atlas de faisceaux et un système de vote (fusion d'étiquettes). D'abord, la première étape consiste en la ré-implémentation de la méthode de segmentation pour qu'elle soit compatible avec plusieurs entrées et un système de vote [Collins et Pruessner, 2010; Iglesias et Sabuncu, 2015; Lampert et al., 2016; Pipitone et al., 2014]. Il est possible d'utiliser un atlas unique et de simplement varier les paramètres liés à la segmentation (seuil de *clustering*, nombre d'itérations pour *SLR*, type de transformation, *pruning*, etc.). Cette notion de multi-paramètres permet de couvrir plus de variabilité, de façon similaire à l'intuition derrière l'approche multi-atlas. L'utilisation de plusieurs atlas de sources diverses (plusieurs sujets, segmentations manuelles produites par différents experts, différents algorithmes de tractographie, etc.) permettra aussi de couvrir un maximum de variabilité. Et les résultats finaux seront obtenus par un vote, encore une fois plusieurs options sont possibles comme le vote majoritaire ou le vote pondéré par une métrique liée à la similarité de forme.

Ensuite, il faut considérer la création de modèles optimaux et trouver les combinaisons optimales de paramètres pour chaque faisceau d'intérêt. Pour qu'un algorithme multi-atlas soit intéressant, il faut qu'il puisse être adapté à plusieurs structures différentes, mais aussi qu'il ne nécessite pas une quantité prohibitive de modèles. L'influence du choix d'atlas (individu sain versus pathologique, âge, nombre de modèles, etc.) et l'optimisation des hyper-paramètres, spécifiques à certains faisceaux, doivent être investiguées soigneusement.

# 4.2 Présentation de l’algorithme Recobundles

Cette section détaille rapidement les étapes principales de l’algorithme *Recobundles* tel que présenté dans la publication [Garyfallidis et al., 2018].

## Recalage global

La première étape consiste à aligner les modèles de faisceaux avec le tractogramme à segmenter. Pour comparer la position spatiale et la forme des modèles aux *streamlines* il est nécessaire d’être dans le même espace. Cette opération est typique des algorithmes classiques de segmentation par atlas. Dans la publication originale, un algorithme basé sur les *streamlines* est utilisé, *Streamlines Linear Registration* [Garyfallidis et al., 2015]. Par souci de robustesse aux jeux de données pathologiques, le recalage linéaire est effectué avec ANTs [Avants et al., 2008] et ensuite appliqué aux *streamlines*.

## Groupement (*Clustering*)

Conceptuellement, il est possible de décrire *Recobundles* comme une méthode comparant les *streamlines* d’un modèle de faisceaux aux *streamlines* d’un tractogramme complet. Par contre, en considérant qu’un tractogramme contient plusieurs millions de *streamlines* et que les modèles en contiennent plusieurs milliers, chaque *streamline* ayant plus d’une dizaine de points, il est impensable de faire une comparaison exhaustive. Pour simplifier la comparaison, les *streamlines* sont groupées à l’aide d’un algorithme de *clustering* basé sur la position et la forme [Garyfallidis et al., 2012, 2016]. Cette étape réduit le tractogramme à quelques milliers de *clusters* et les modèles à quelques centaines de *clusters*. L’utilisation des centroïdes, une *streamline* moyenne, de ces *clusters* permet de grandement accélérer les opérations de comparaison.



## 4.2. PRÉSENTATION DE L’ALGORITHME RECOBUNDLES

### Réduction de l’espace de recherche

Une comparaison exhaustive est ensuite effectuée entre les centroïdes provenant des *clusters* du tractogramme et ceux d’un des modèles. Un seuil de similarité relativement élevé est ensuite appliqué, ce qui permet d’éliminer la vaste majorité des *clusters* pour le restant de l’algorithme. Ces *clusters* sont considérés comme complètement différents du modèle en cours et tout calcul subséquent peut les ignorer.

### Recalage local

Considérant la structure des faisceaux de la matière blanche et le fait qu’un algorithme de recalage d’image (e.g ANTs) n’en prend pas compte, un recalage local dans l’espace des *streamlines* entre le voisinage, précédemment calculé, et le modèle est requis [Garyfallidis et al., 2015]. L’objectif est de minimiser la distance entre le modèle et le voisinage en utilisant les *streamlines* comme structure plutôt que les intensités des niveaux de gris d’une image. Cette opération permet de rendre l’algorithme plus robuste aux déformations locales dues aux pathologies ou simplement dues à la variabilité anatomique.

### Élagage (Pruning)

Une fois que le modèle est positionné correctement dans l’espace du tractogramme (réduit), un second seuil de similarité, plus conservateur cette fois, est appliqué. Ce seuil représente la tolérance de distance entre le modèle et ce qui sera segmenté. Comme mentionné plus tôt, ces opérations sont effectuées sur les centroïdes et non sur les *streamlines*. L’étape suivante est celle de *pruning*, les *clusters* considérés suffisamment similaires au modèle permettent de classifier les *streamlines* qui leur sont associées. Le résultat final est un sous-ensemble de *streamlines* du tractogramme initial similaire au modèle.

## 4.3 Multi-Atlas et Multi-paramètres

### 4.3.1 Choix d’atlas

Le choix d’atlas est l’une des principales difficultés liées à l’algorithme, non seulement dû aux choix complexes que nécessite la définition de faisceaux en neuroanatomie, mais aussi dû à l’étape simpliste de sélection de *streamlines* de *Recobundles*. Toutes les *streamlines* aberrantes présentes dans le modèle seront considérées durant la segmentation et vont potentiellement inclure des résultats non désirés. Cette limitation est l’une des raisons justifiant une approche multi-atlas. En utilisant plusieurs modèles à la fois les *streamlines* aberrantes vont varier et ainsi la probabilité d’être sélectionné par une majorité des exécutions sera ainsi réduite. De plus, comme les modèles doivent être construits à partir de segmentations manuelles (contenant de la variabilité, voir Chapitre 3), cette variabilité peut être prise en compte par le système de vote. En effet, un vote plus ou moins strict permettra de contrôler le degré d’accord entre tous les modèles.

Tel qu’originellement proposé et évalué dans la publication de Garyfallidis et al. [2018], l’algorithme utilisait un seul modèle provenant d’un expert pour segmenter un tractogramme. En sachant, grâce au résultat de Tractostorm (Chapitre 3), que les segmentations d’experts possèdent une variabilité inhérente (avec soi-même et entre experts) il est difficile de justifier l’utilisation d’un seul modèle. Il est aussi injuste d’évaluer l’algorithme en comparant le résultat à un seul expert et de considérer celui-ci comme «*parfait*».

### 4.3.2 Paramètres optimaux

De nombreux paramètres sont à choisir et leur influence est majeure, cette problématique est aussi directement adressée par l’approche de fusion d’étiquettes. En choisissant un ensemble de paramètres acceptables, la probabilité d’inclure un artefact méthodologique dû aux paramètres est réduite. L’algorithme original ne permet pas l’exploration de plusieurs paramètres, ce qui rendait les résultats extrêmement variables (voir Section 4.5). De plus, grâce à l’approche multi-paramètres la généralisation à des jeux de données variées (pathologie, population âgée) et à plusieurs

### 4.3. MULTI-ATLAS ET MULTI-PARAMÈTRES

faisceaux de forme parfois très différente est simplifiée. L'ensemble de paramètres permet d'explorer l'espace de solution et la fusion d'étiquettes permet d'éviter les aberrations.

#### 4.3.3 Fusion d'étiquettes

Une fois l'algorithme *Recobundles* exécuté plusieurs fois, il est nécessaire d'agréger tous les résultats. Chaque exécution a sélectionné un ensemble différent de *streamlines* pour chaque faisceau. Dans une situation où 5 atlas de 20 faisceaux sont utilisés avec 10 jeux de paramètres, chaque *streamline* aura 20 scores (chaque faisceau) de 0 à 50 (5 atlas \* 10 paramètres). Pour qu'une *streamline* soit classifiée comme appartenant à un faisceau il est important qu'elle ait le plus haut score parmi tous les faisceaux. En effet, pour faciliter la distinction entre les faisceaux, une *streamline* ne peut faire partie de plus d'un faisceau à la fois.

Ensuite, selon le degré de certitude spécifié par l'utilisateur un score minimum est nécessaire. Par exemple, un score minimum de 50/50 revient à faire l'intersection (opération sur les ensembles) entre tous les éléments de chaque exécution. Un score minimum de 1/50 revient à faire l'union entre tous les éléments de chaque exécution. Par contre, ces deux seuils sont très risqués étant donné qu'une seule exécution aberrante peut facilement miner la qualité du résultat final. En utilisant des seuils variant entre 10/20 ou 30/40 il est facile d'obtenir un compromis entre une segmentation libérale ou conservatrice tout en éliminant les *streamlines* résultant d'une exécution aberrante. La Figure 3.6 au Chapitre précédent montre l'effet d'une transition entre vote libéral vers vote conservateur. L'utilisation de *Recobundles* avec plusieurs atlas, plusieurs paramètres et la méthode de fusion d'étiquettes sera dorénavant nommée *RecobundlesX*.

## 4.4 Évaluation et métriques

Cette section détaille rapidement les métriques existantes pour la comparaison de faisceaux, entre autres, celles utilisées pour l'évaluation des résultats. Certains de ces détails ont déjà été mentionnés au Chapitre 3.

### 4.4.1 Dice similarity coefficient

Dans les situations où la reproductibilité d'un algorithme doit être étudiée, par exemple entre deux reconstructions d'un même faisceau ou la segmentation d'un même faisceau par deux experts, le *Dice similarity coefficient*, ou coefficient Dice, [Dice, 1945] est une mesure simple et bien connue. Lorsqu'appliquée avec des données de tractographie sous forme de faisceaux, l'opération consiste à binariser les voxels traversés par les *streamlines* pour ensuite calculer leur chevauchement à l'aide de l'Équation 4.1 suivante. Cette métrique et ses variantes produisent un chiffre entre 0 et 1, où 0 ne représente aucun chevauchement et 1 représente un chevauchement parfait.

$$D(B_1, B_2) = \frac{2 * \sum_v (B_{1,v} \cap B_{2,v})}{\sum_v (B_{1,v}) + \sum_v (B_{2,v})} \quad (4.1)$$

#### Weighted-Dice

La version pondérée par la densité du coefficient Dice a été introduite par Cousineau et al. [2017] dans le but de diminuer l'impact des *streamlines* aberrantes. En effet, dans la version binaire quelques dizaines de *streamlines* s'écartant de la norme peuvent avoir autant d'influence que plusieurs milliers de *streamlines* passant par les mêmes voxels. Cette métrique représente plutôt le chevauchement entre la partie centrale (cœur) d'un faisceau et celle-ci se calcule à l'aide d'une carte de densité plutôt qu'une carte binaire.

$$WD(B_1, B_2) = \frac{\sum_{v'} (B_{1,v}) + \sum_{v'} (B_{2,v})}{\sum_v (B_{1,v}) + \sum_v (B_{2,v})} \quad (4.2)$$

## 4.4. ÉVALUATION ET MÉTRIQUES

### Streamline-wise Dice

Dans le cas où une segmentation d'un même tractogramme est accomplie plus d'une fois, et donc que les *streamlines* sont identiques, il est possible d'étendre la notion du coefficient Dice aux indices qui représentent chaque *streamline*. En effet, plutôt que d'utiliser les voxels pour mesurer le chevauchement il est possible de comparer les *streamlines* elles-mêmes et donc de voir si deux segmentations possèdent des *streamlines* en commun. La formule est identique au coefficient Dice standard, mais les entités comparées sont les indices représentant les *streamlines*, plutôt que les voxels traversés par celles-ci.

### 4.4.2 Bundle adjacency

La *Bundle adjacency* (BA) permet d'évaluer la similarité entre deux faisceaux. Les faisceaux peuvent être représentés soit sous la forme de *streamlines* ou sous la forme de carte binaire voxelisée. Il est important de noter que contrairement aux métriques présentées ci-dessus, la BA ne donne pas une valeur entre 0 et 1, mais bien une valeur en millimètre.

### Voxel-wise bundle adjacency

Cette métrique permet de quantifier la distance entre deux faisceaux. La première étape consiste à convertir les *streamlines* de chaque faisceau en masque binaire dans une grille de la même résolution que les données originales (1mm isotropique). Ensuite pour chaque voxel d'un masque binaire qui n'est pas en superposition avec son analogue, le plus proche voisin dans l'analogue est trouvé et la distance euclidienne calculée. Ceci permet d'estimer la distance à la superposition et ainsi d'avoir une approximation de la distribution spatiale des données aberrantes. Cette valeur prise en considération avec le voxel-wise Dice est particulièrement utile et facile à interpréter.

#### 4.4. ÉVALUATION ET MÉTRIQUES

##### Streamline-wise bundle adjacency

Cette métrique mesure le voisinage entre les *clusters* présents dans les faisceaux Garyfallidis et al. [2015]. En effet, la première étape de cette méthode est de grouper les *streamlines* de chaque faisceau Garyfallidis et al. [2012], ce qui réduit la charge computationnelle. Ensuite, une comparaison par paires des *clusters* est effectuée à l'aide de la distance MDF Garyfallidis et al. [2010], ce qui produit une matrice NxM où N et M sont le nombre de *clusters* dans chaque faisceau. Finalement, pour obtenir une mesure de proximité entre les deux faisceaux il est nécessaire de trouver la valeur minimale de distance (présente dans la matrice) pour chaque *cluster* pour chacun des deux faisceaux (voir les Équations 4.3). Cette mesure de similarité inclut une information sur la forme, contrairement à la version voxelique qui ne contenait qu'une information spatiale. La Figure 4.1 permet de visualiser les étapes d'un calcul de BA ainsi que le résultat obtenu. La BA représente la distance moyenne entre les *clusters* et leur plus proche voisin. Cette valeur prise en considération avec le streamline-wise Dice est particulièrement utile et facile à interpréter.

$$\begin{aligned} avg\_closest\_cluster(B_1, B_2) &= avg(min(MDF\_dist\_matrix(B_1, B_2), axis = 0)) \\ BA(B_1, B_2) &= \frac{avg\_closest\_cluster(B_1, B_2) + avg\_closest\_cluster(B_2, B_1)}{2} \end{aligned} \quad (4.3)$$

#### 4.4. ÉVALUATION ET MÉTRIQUES

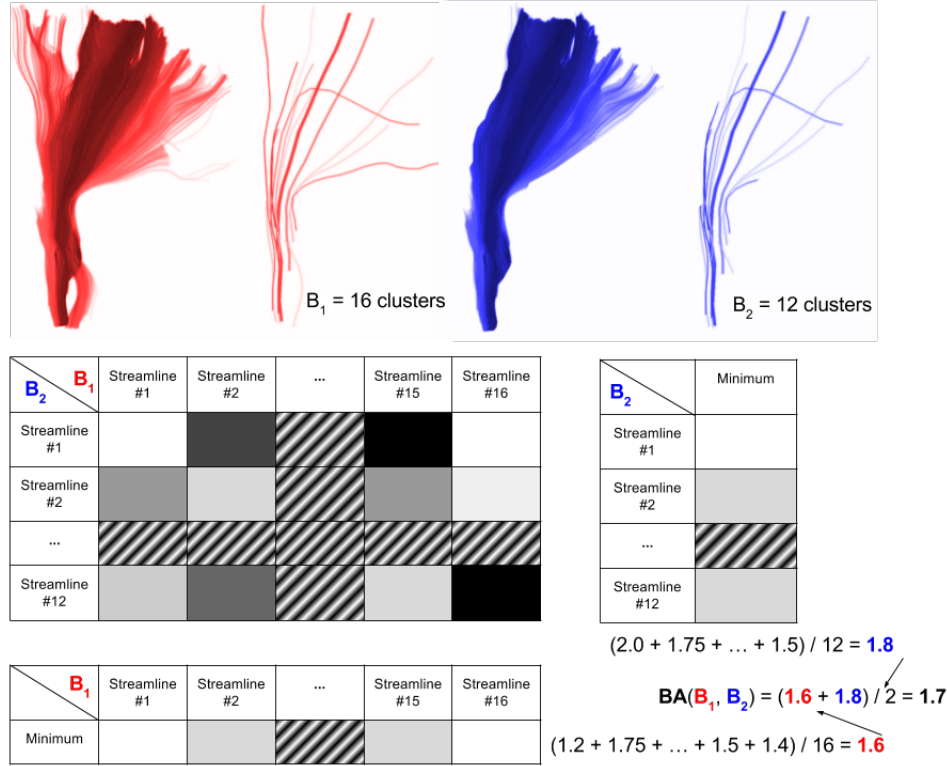


figure 4.1 – Le calcul de BA commence avec un ensemble de centroïdes provenant de l’algorithme QuickBundle [Garyfallidis et al., 2012], la matrice d’adjacence est calculée, les valeurs représentent la distance MDF entre chaque paire de centroïdes. La distance minimale de l’ensemble de centroïdes #1 à l’ensemble de centroïdes #2 est obtenue en faisant la moyenne des valeurs minimales par colonnes. La même opération est faite pour les rangées afin d’obtenir la relation entre les centroïdes des ensembles #2 à #1 (non-symétrique).

#### 4.4.3 Classification binaire

Les métriques de classification binaire permettent de comparer des données à des vérités terrain, dans la situation où un algorithme de segmentation est comparé à un expert ou un consensus composé de plusieurs experts il est possible de compter l’expertise humaine comme une valeur fiable. Ces métriques se basent sur des calculs faits à partir de mesures de vrais positifs, faux positifs, vrais négatifs et faux négatifs. La Figure 4.2 représente ces métriques et comment les obtenir.

#### 4.4. ÉVALUATION ET MÉTRIQUES

Ces opérations sur des ensembles sont les mêmes, peu importe la représentation, voxels ou *streamlines*. Dans la présente situation, une segmentation d'un observateur (voxels ou *streamlines*) contiendra des éléments en commun avec la vérité terrain qui seront nommés *vrais positifs*. Les éléments sélectionnés par l'observateur qui ne seront pas dans la vérité terrain seront des *faux positifs*. Les éléments omis dans les deux ensembles sont des *vrais négatifs* et ceux omis par l'observateur, mais présents dans la vérité terrain seront des *faux négatifs*.

#### Kappa & Dice

Les coefficients de Dice et Kappa mesurent le degré de concordance entre deux évaluateurs. Le coefficient de Kappa permet de quantifier la probabilité d'un accord se produisant par chance. Le coefficient Dice a été mentionné dans la Section 4.4.1, en effet cette mesure peut être utilisée entre deux évaluateurs ou entre un évaluateur et une vérité terrain. L'utilisation dans les deux contextes est identique, mais l'interprétation est différente.

Considérant la dimensionnalité de nos données, c'est-à-dire qu'en ayant des millions de *streamlines* ou des millions de voxels, les coefficients de Dice et Kappa sont très similaires. En effet, étant donné que la probabilité d'avoir un accord sur quelques milliers d'éléments dans un ensemble de plus d'un million est extrêmement basse.

$$Dice = \frac{2 * TP}{2 * TP + FP + FN} \quad (4.4)$$

$$Kappa = 1 - \frac{1 - p_o}{1 - p_e} \quad (4.5)$$

$$p_o = \frac{TP+TN}{P+N}$$
$$p_e = \frac{1}{N^2} \sum_k n_{k1} * n_{k2}$$



#### 4.4. ÉVALUATION ET MÉTRIQUES

##### Sensitivité et Spécificité

La sensibilité et spécificité sont des mesures très courantes et ont une importance majeure pour résumer les performances d'un test de détection. La sensibilité mesure la capacité à capturer les vrais positifs, tandis que la spécificité mesure la capacité à capturer les vrais négatifs. Ces deux mesures doivent être interprétées ensemble pour obtenir une interprétation utile. En effet, une méthode avec une sensibilité 99% ne veut rien dire si on ne connaît pas sa spécificité. Si la spécificité est de 1%, cela voudrait dire que la méthode ne fait que tout classer comme positif et donc n'a aucune utilité réelle. Par contre, si la spécificité est de 99% cela veut dire que la méthode classe correctement la vaste majorité des éléments.

$$Sensitivity = \frac{TP}{TP + FN} \quad (4.6)$$

$$Specificity = \frac{TN}{TN + FP} \quad (4.7)$$

##### Précision et exactitude

Lorsque que les valeurs de sensibilité et de spécificité ne sont pas suffisamment indicatives de la situation (e.g lorsqu'elles sont entre 30% et 70%), il est nécessaire d'ajouter d'autres mesures pour améliorer l'interprétation d'une situation. La précision permet de mesurer la fraction d'éléments sélectionnés qui sont utiles à la classification positive. L'exactitude permet de mesurer la proportion de vrais résultats, qu'ils soient positifs ou négatifs, sur l'ensemble d'éléments en entier. Pour faciliter la présentation des résultats, uniquement les mesures de similarité et de classification binaire liées à la représentation voxelique seront présentées.

$$Precision = \frac{TP}{TP + FN} \quad (4.8)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (4.9)$$

## 4.5. EXPÉRIMENTATIONS

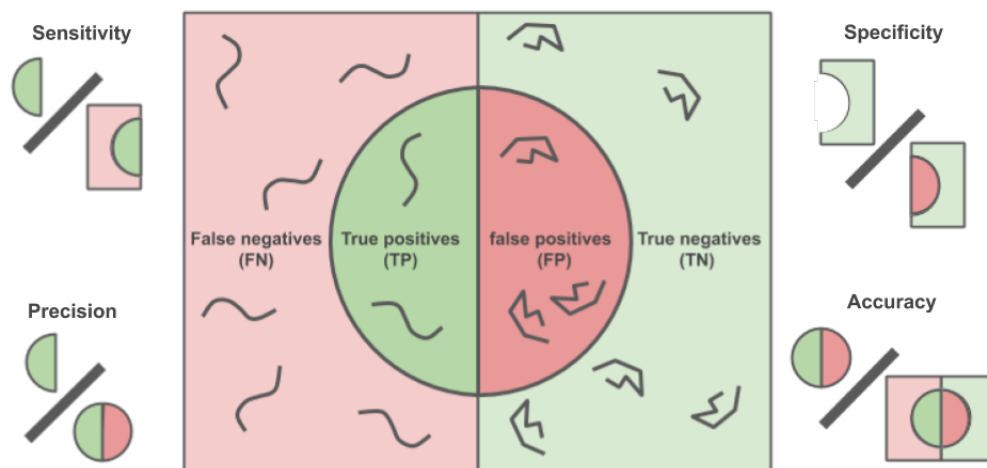


figure 4.2 – Les éléments à l’intérieur du rectangle entier représentent tout ce qui doit être classifié. Les lignes courbes (petit rectangle à gauche) sont des éléments «positifs», tandis que celles brisées (petit rectangle à droite) sont des éléments «négatifs». La coloration rouge représente des éléments mal classifiés et la coloration verte représente des éléments correctement classifiés. Le cercle (rouge et vert) représente la sélection d’un observateur tentant une classification.

## 4.5 Expérimentations

### 4.5.1 Jeu de données

L’évaluation de l’algorithme a été faite à partir des données issues du projet Tractostorm décrit à la Section 3. Le jeu de données est composé des segmentations du faisceau pyramidal (PyT) de 5 sujets du Human Connectome Project (HCP) effectuées par 24 participants. Dans le cadre du projet Tractostorm chaque sujet HCP a été tripliqué, ceci veut dire qu’au total environ 360 exemplaires du PyT étaient disponibles. C’est à partir de ces 360 exemplaires que l’ensemble d’entrées *multi-atlas* provient.

Dans le cadre du même projet, un «*gold standard*», pour chaque sujet HCP a été généré à partir de la moyenne des segmentations provenant du sous-groupe ayant une expertise en neuroanatomie (groupe expert). Ces mêmes «*gold standards*» seront considérés comme la vérité terrain pendant l’évaluation de l’algorithme, plus de détails sur la génération de ces vérités terrain sont disponibles dans la Section 3.2.4.

## 4.5. EXPÉRIMENTATIONS

En d’autres mots l’objectif est d’utiliser un sous-ensemble des segmentations pris au hasard pour segmenter le PyT et comparer ce résultat à la segmentation moyenne des «*experts*».

### 4.5.2 Ensemble de configurations

L’algorithme peut aussi prendre en entrée plusieurs paramètres et ainsi exécuter l’algorithme plusieurs fois avec des paramètres différents pour mieux explorer l’espace des possibilités. Les 3 paramètres pouvant varier sont le seuil de groupement du tractogramme (*whole brain clustering threshold*), le seuil de groupement des modèles (*models clustering threshold*) et le seuil pour l’élagage (*pruning threshold*). Chaque paramètre avait 5 possibilités, ce qui mène à 125 possibilités de triplet de paramètres. Ces paramètres sont enregistrés dans un fichier de configuration et sélectionnés au hasard par l’algorithme dépendamment du nombre d’exécutions demandé. Il est important de mentionner que le nombre d’exécutions lié au nombre de jeu de paramètres se fait de façon multiplicative au nombre d’atlas. C’est-à-dire que si 8 atlas sont utilisés comme entrée et que 16 jeux de paramètres sont demandés, l’algorithme *Recobundles* sous-jacent sera exécuté 128 fois.

### 4.5.3 Exécutions de l’algorithme

Étant donné l’ensemble gigantesque des possibilités d’atlas et de paramètres, seulement un sous-ensemble de ces combinaisons a pu être testé. 10 groupes de 1, 2, 4, 8, 16 combinaisons d’atlas (50) ont été sélectionnés, ainsi que 10 jeux de 1, 2, 4, 8, 16 combinaisons de paramètres (50) et pour chacun des sujets HCP en entrée. Ceci cumule à 12 500 commandes au total pour *RecobundlesX*, chaque commande pouvant varier de 2 minutes (1 atlas et 1 jeu de paramètres) à plus de 60 minutes (16 atlas et 16 jeux de paramètres). La fusion d’étiquettes a été effectuée à l’aide d’un vote majoritaire (50%) sur la représentation de *streamlines*. Ainsi, si une commande est lancée à 1 atlas et 1 jeu de paramètres (1 exécution total), elle sera en fait équivalente à *Recobundles* et dès que le nombre d’exécutions total augmente il s’agit de *RecobundlesX*. La Figure 4.3 illustre comment les groupes et les commandes sont construits.

## 4.6. RÉSULTATS

<b>Sujets:</b>	$S_1, S_2, S_3, S_4, S_5$
<b>Atlas:</b>	$A_{1a}, A_{1b}, A_{1c}, A_{1d}, A_{1e}, A_{1f}, A_{1g}, A_{1h}, A_{1i}, A_{1j}$ $A_{2a}, A_{2b}, A_{2c}, A_{2d}, A_{2e}, A_{2f}, A_{2g}, A_{2h}, A_{2i}, A_{2j}$ $A_{4a}, A_{4b}, A_{4c}, A_{4d}, A_{4e}, A_{4f}, A_{4g}, A_{4h}, A_{4i}, A_{4j}$ $A_{8a}, A_{8b}, A_{8c}, A_{8d}, A_{8e}, A_{8f}, A_{8g}, A_{8h}, A_{8i}, A_{8j}$ $A_{16a}, A_{16b}, A_{16c}, A_{16d}, A_{16e}, A_{16f}, A_{16g}, A_{16h}, A_{16i}, A_{16j}$
<b>Paramètres:</b>	$P_{1a}, P_{1b}, P_{1c}, P_{1d}, P_{1e}, P_{1f}, P_{1g}, P_{1h}, P_{1i}, P_{1j}$ $P_{2a}, P_{2b}, P_{2c}, P_{2d}, P_{2e}, P_{2f}, P_{2g}, P_{2h}, P_{2i}, P_{2j}$ $P_{4a}, P_{4b}, P_{4c}, P_{4d}, P_{4e}, P_{4f}, P_{4g}, P_{4h}, P_{4i}, P_{4j}$ $P_{8a}, P_{8b}, P_{8c}, P_{8d}, P_{8e}, P_{8f}, P_{8g}, P_{8h}, P_{8i}, P_{8j}$ $P_{16a}, P_{16b}, P_{16c}, P_{16d}, P_{16e}, P_{16f}, P_{16g}, P_{16h}, P_{16i}, P_{16j}$

figure 4.3 – Représentation des commandes générées pour *RecobundlesX*. Les groupes d’atlas ou de paramètres sont aléatoirement construits. Par exemple, un total de  $2.712264755 \times 10^{27}$  ou  $\binom{360}{16}$  possibilités est accessible pour la construction d’un groupe de 16 atlas, mais seulement 10 groupes dans cet ensemble sont «*pigés*». Pour chaque commande, un sujet, un groupe d’atlas et un groupe de paramètres sont sélectionnés. Par exemple, une commande pour  $S_1$  avec  $A_{16a}$  et  $P_{8b}$  résultera en une segmentation du premier sujet avec 128 exécutions. Toutes les combinaisons de ces 3 facteurs sont explorées ( $5 * 50 * 50 = 12\ 500$ ).

## 4.6 Résultats

Afin de présenter les résultats de manière plus intuitive, seule la combinaison des exécutions venant des atlas et des jeux de paramètres sera rapportée. C’est-à-dire qu’uniquement les nombres totaux d’exécutions (nombre d’atlas multiplié par nombre de jeu de paramètres) seront présentés. L’influence de ces facteurs et les cas limites de cette approche sont discutés plus en détail dans la [Discussions](#). La Figure 4.4 montre comment *Recobundles* peut être variable et comment *RecobundlesX* peut «*stabiliser*» les résultats. Le résultat de l’algorithme *RecobundlesX* est un faisceau composé de *streamlines*. Celui-ci possède des mesures de longueur et, lorsque discrétisés sur une grille, un volume qui peuvent être rapportés. De la même manière que dans le Chapitre 3, ces mesures ne sont pas directement en lien avec la qualité de la segmentation en comparaison à la vérité terrain, mais ces mesures et leurs variabilités peuvent procurer des informations utiles sur les résultats.

## 4.6. RÉSULTATS

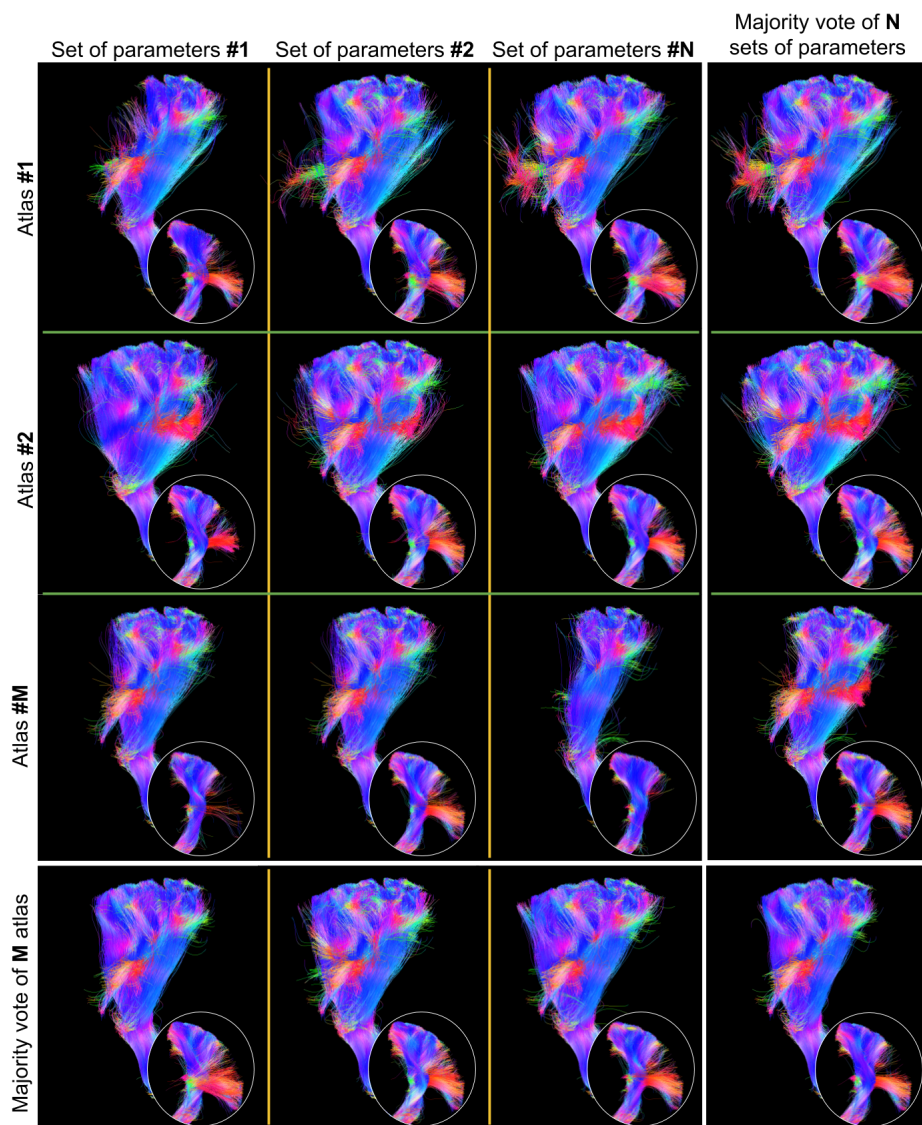


figure 4.4 – Comparaison visuelle des résultats individuels de *Recobundles* et *RecobundlesX* par jeu de paramètres (rangées) et par atlas (colonnes). Chaque élément de la grille représente le résultat d’une exécution avec une vue sagittale et en vignette un zoom de la vue coronale. La dernière colonne représente le résultat de *RecobundlesX* assemblant les exécutions multi-paramètres. La dernière rangée représente le résultat de *RecobundlesX* assemblant les exécutions multi-atlas. L’élément au coin inférieur gauche représente le résultat de *RecobundlesX* assemblant les exécutions multi-paramètres et multi-atlas. La fusion d’étiquettes se fait toujours à l’aide d’un vote majoritaire. Dans le cas de la présente étude, les valeurs maximales sont  $N = 16$  et  $M = 16$ , la figure a été réduite pour des fins de simplification.

## 4.6. RÉSULTATS

	Average length (mm)		Volume (cm3)		Streamline count	
# runs	MED	IQR	MED	IQR	MED	IQR
1	137.67	8.44	69.63	21.15	12282	10477
2	136.85	7.48	81.55	19.29	17919	14773
4	138.29	7.88	76.08	17.49	17000	12888
8	137.54	8.87	72.39	21.12	16476	14184
16	136.29	8.23	71.10	22.10	16437	14353
32	136.04	8.02	68.31	16.62	14933	13527
64	137.59	7.34	65.98	10.37	14365	10644
128	137.04	5.66	64.95	9.93	12718	10384
256	137.32	5.79	65.58	10.08	13566	10202

tableau 4.1 – Tableau présentant les résultats agrégés pour les exécutions multi-atlas ainsi que multi-paramètres. Ces mesures ne sont pas directement indicatives de la variabilité ou de la similarité avec la vérité terrain, ces mesures sont typiquement celles rapportées dans le contexte d’une étude.

Ce qui est particulièrement intéressant de ces mesures est la variance des résultats. Il est surprenant de voir à quel point les résultats sont variables, ultimement avec l’utilisation de *Recobundles* ou *RecobundlesX* tous ces résultats peuvent être considérés valides. Par contre, on peut observer dans la Figure 4.5 que le nombre de résultats aberrants diminue avec le nombre d’exécutions total. Une certaine augmentation de la variabilité est observable entre 2 et 16 exécutions total, mais cette variabilité semble redescendre par la suite. Les données sur la longueur moyenne des *streamlines* sont très stables, cette mesure est rarement rapportée dans le contexte d’étude scientifique. Par contre, le nombre de *streamlines* et le volume sont des mesures couramment rapportées et souvent considérées comme des variables cruciales en tractographie.

## 4.6. RÉSULTATS

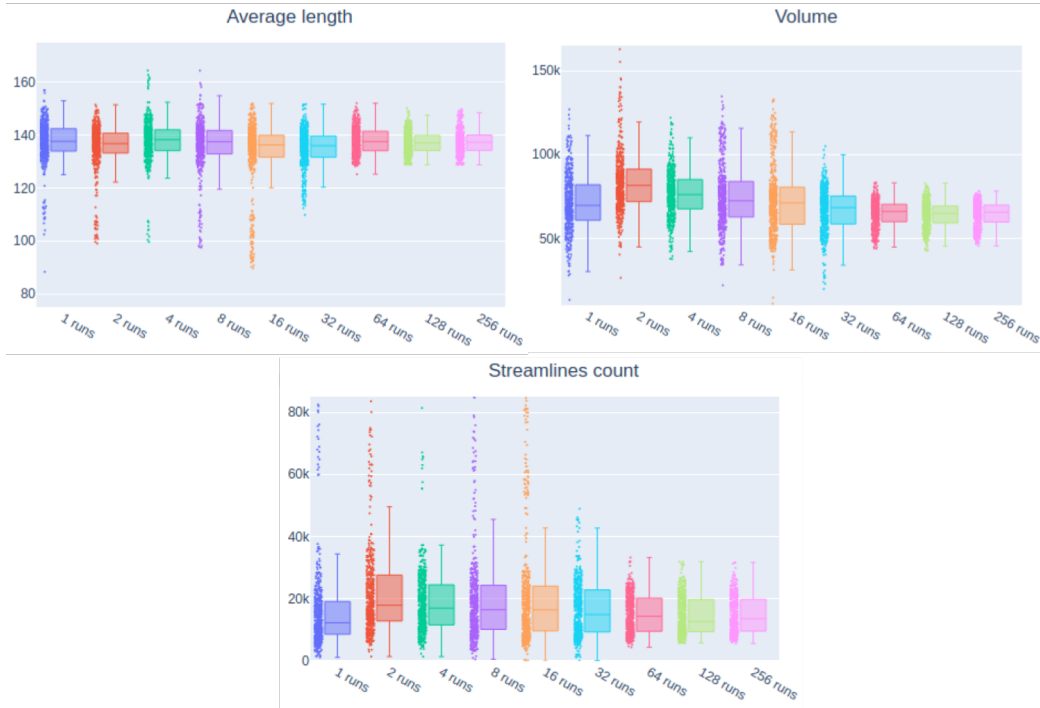


figure 4.5 – Graphique en boîte à moustaches (*box-and-whisker plot*) présentant les mesures individuelles, la diminution de résultats aberrants permet la réduction de l'intervalle interquartile, tout en conservant une médiane similaire.

Il est intéressant que les résultats de *RecobundlesX* résident dans un intervalle chevauchant les segmentations manuelles issues de Tractostorm. Comme mentionné plus tôt, ces mesures n'indiquent en rien la similarité entre les résultats et la vérité terrain. En effet, deux faisceaux complètement différents peuvent avoir le même volume ou le même nombre de *streamlines*. Pour véritablement comparer les résultats, il est important de considérer les mesures présentées à la Section 4.4.

Le Tableau 4.2 et Figure 4.6 montrent le chevauchement (coefficient Dice) ainsi que la distance moyenne (bundle adjacency) entre les résultats et les vérités terrain. Ces mesures permettent de réellement évaluer la similarité et l'exactitude des résultats. Pour des fins de simplicité, uniquement la représentation par voxels est présentée, mais la représentation par *streamlines* possède des tendances similaires. Ici un score de 1 pour le coefficient Dice et 0 pour la *Bundle adjacency* indiquerait une segmentation identique à la vérité terrain.

## 4.6. RÉSULTATS

# runs	Dice Coefficient		Bundle adjacency (mm)	
	MED	IQR	MED	IQR
1	0.70	0.08	2.81	1.13
2	0.67	0.08	2.89	0.95
4	0.70	0.07	2.70	0.84
8	0.71	0.08	2.65	0.89
16	0.72	0.08	2.67	0.96
32	0.75	0.07	2.44	0.83
64	0.76	0.05	2.36	0.65
128	0.77	0.04	2.31	0.60
256	0.77	0.04	2.35	0.58

tableau 4.2 – Mesure de chevauchement et de distance dans la représentation voxélique. La tendance observée indique une amélioration de l'accord (augmentation du coefficient Dice et diminution de la BA) et une réduction de la variabilité à mesure que le nombre d'exécutions augmente. Cette tendance est également observable pour la représentation par *streamlines*.

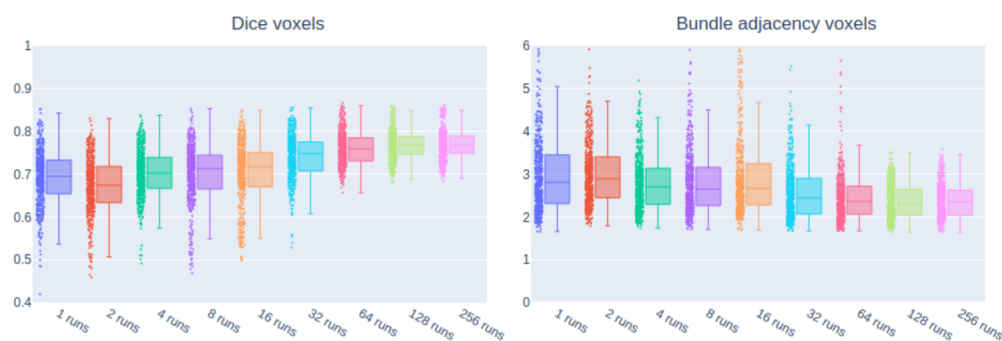


figure 4.6 – Comme observé pour les mesures individuelles, les données aberrantes sont réduites ce qui diminue l'intervalle interquartile. De plus, les mesures d'accord s'améliorent à mesure que le nombre d'exécutions augmente.



## 4.6. RÉSULTATS

Deux choses sont intéressantes à noter, l'amélioration (augmentation du coefficient Dice et diminution de la *Bundle Adjacency*) de la médiane à mesure que le nombre d'exécutions total augmente et la diminution de l'IQR. La réduction dans les données aberrantes diminue aussi avec le nombre d'exécutions total.

Le concept de chevauchement n'a pas suffisamment de nuance pour indiquer la capacité de l'algorithme à identifier les faux et vrais positifs ainsi que les faux et vrais négatifs. Donc, les mesures de classifications binaires sont nécessaires, encore une fois uniquement la représentation par voxels est présentée, mais la représentation par *streamlines* possède des tendances similaires.

# runs	Sensitivity		Specificity		Accuracy		Precision	
	MED	IQR	MED	IQR	MED	IQR	MED	IQR
1	0.88	0.12	0.91	0.06	0.90	0.04	0.66	0.13
2	0.92	0.08	0.88	0.06	0.88	0.04	0.60	0.11
4	0.91	0.09	0.90	0.05	0.90	0.03	0.64	0.11
8	0.91	0.12	0.91	0.06	0.91	0.04	0.67	0.13
16	0.89	0.14	0.92	0.05	0.91	0.03	0.68	0.13
32	0.90	0.11	0.93	0.05	0.92	0.02	0.72	0.12
64	0.89	0.09	0.94	0.03	0.93	0.01	0.74	0.09
128	0.90	0.06	0.94	0.02	0.93	0.01	0.74	0.07
256	0.91	0.08	0.95	0.03	0.93	0.01	0.74	0.08

tableau 4.3 – La mesure de classification binaire pour la représentation voxélique. On peut observer une tendance où la médiane (de toutes les mesures) augmente légèrement et la variabilité diminue. Typiquement, les algorithmes balancent entre augmentation de la sensibilité et diminution de la spécificité (ou l'inverse), par contre dans le cas présent les deux mesures augmentent. La diminution de l'intervalle inter-quartile indique une stabilisation des résultats vers un meilleur chevauchement avec la vérité terrain.

## 4.6. RÉSULTATS

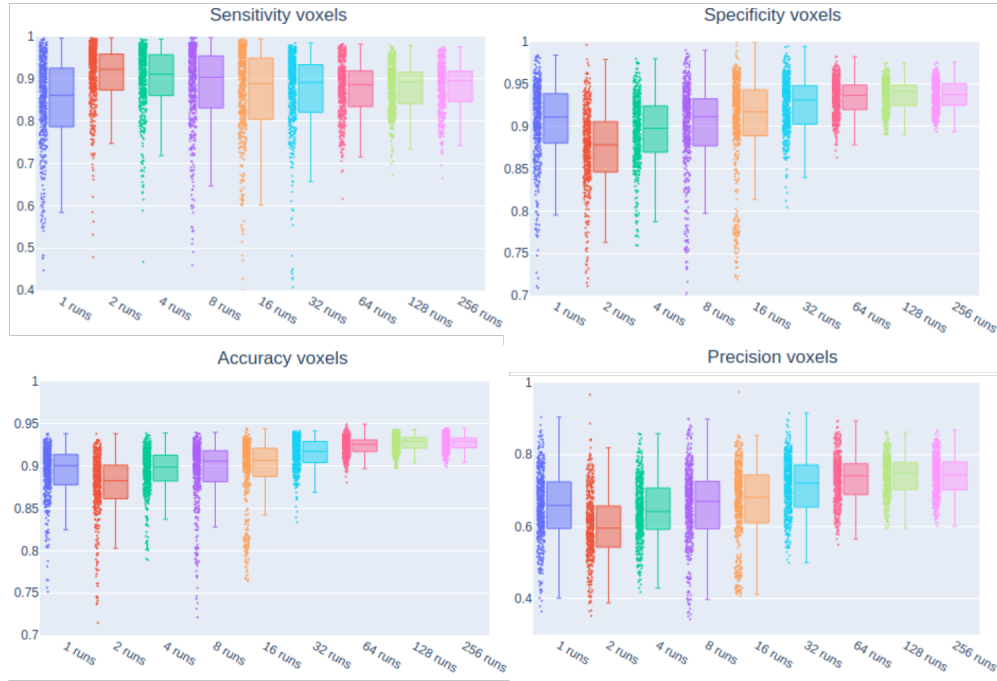


figure 4.7 – Malgré la légèreté de l'amélioration de la médiane, la quantité de résultats aberrants disparus ne semble qu'avoir une influence mineure sur l'intervalle interquartile. Seulement une minorité des exécutions mène à des résultats aberrants. Globalement ceci indique qu'un plus haut niveau de confiance peut être accordé aux résultats de *RecobundlesX* à mesure que le nombre d'exécutions augmente.

Encore une fois deux tendances sont intéressantes à noter pour toutes les mesures de classification binaire, l'augmentation de la médiane à mesure que le nombre d'exécutions total augmente et la diminution de l'IQR. L'augmentation de la médiane est plutôt faible considérant le nombre d'exécutions total, par contre il est très rare que la sensibilité et la spécificité augmentent ensemble. Cela veut dire que l'algorithme classe mieux les vrais positifs ainsi que les vrais négatifs. L'augmentation de la précision indique que les segmentations contiennent moins de faux positifs. L'augmentation de la médiane pour toutes les mesures de classification binaire et la diminution de leur IQR indiquent que les résultats peuvent être considérés comme supérieurs à mesure que le nombre total d'exécutions augmente.

## 4.6. RÉSULTATS

### 4.6.1 Atlas basés sur les vérités terrain

Plutôt que de «*piger*» dans l'ensemble des segmentations manuelles du projet Tractostorm, il est possiblement plus judicieux de sélectionner les vérités terrain issues de 4 sujets HCP pour segmenter un 5e sujet HCP (validations croisées). Ainsi, des entrées aberrantes peuvent être évitées et grandement accélérer le temps d'exécution en limitant à 4 atlas de qualité, plutôt que 16 atlas (ou plus) de qualité variable.

De façon similaire à la sélection aléatoire d'atlas, les mesures telles que la longueur moyenne, volume et nombre de *streamlines* n'indiquent en rien la similarité à la vérité terrain. Par contre, la Table 4.4 montre que l'écart interquartile (IQR) diminue avec le nombre d'exécutions.

	Average length (mm)		Volume (cm3)		Streamline count	
# runs	MED	IQR	MED	IQR	MED	IQR
1	137.21	7.85	73.90	16.15	15690	11169
2	137.40	5.69	84.55	16.93	19266	13237
4	138.14	6.14	78.79	15.60	19517	12220
8	139.05	5.88	76.66	14.87	19368	10315
16	138.79	5.79	74.52	9.04	19576	9720
32	138.07	6.15	71.74	8.53	18461	11954
64	138.23	5.89	71.78	8.16	18754	11420

tableau 4.4 – Les résultats des mesures individuelles issues des exécutions basées sur les vérités terrain ne sont pas nécessairement en accord avec celles présentées ci-haut. Sans les mesures d'accord ou de classification binaire, il est impossible de savoir quelle approche est la plus rapprochée de la vérité terrain.

Les mesures de chevauchement et de distance de la Table 4.5 montrent une tendance similaire à l'expérimentation précédente. En effet, à mesure que le nombre d'exécutions augmente la médiane s'améliore pour les 2 mesures et l'IQR diminue. De plus, à un nombre similaire d'exécutions les mesures de chevauchement sont similaires à l'expérimentation précédente. Par contre, la mesure de distance est inférieure ainsi que l'IQR ce qui indique des segmentations relativement plus proches d'un che-

## 4.6. RÉSULTATS

vauchement.

# runs	Dice Coefficient		Bundle adjacency (mm)	
	MED	IQR	MED	IQR
1	0.71	0.07	2.52	0.87
2	0.68	0.06	2.68	0.68
4	0.70	0.07	2.46	0.71
8	0.72	0.06	2.38	0.60
16	0.73	0.05	2.26	0.52
32	0.75	0.06	2.28	0.48
64	0.75	0.05	2.26	0.48

tableau 4.5 – Mesure d’accord pour la représentation voxélique. Le nombre d’exécutions est limité à 64 dû au fait qu’un maximum de 4 atlas peut être utilisé. À nombre d’exécutions égal, la qualité de l’accord est similaire ce qui indique que l’utilisation de modèles aléatoires variés ou de modèles sélectionnés/désignés n’a qu’une influence limitée.

Finalement, les mesures de classification binaire ont encore une fois des tendances similaires, c’est-à-dire une amélioration globale de toutes les mesures et de leur IQR. Par contre, la sensibilité et spécificité sont respectivement légèrement supérieure et inférieure, ce qui fait diminuer l’exactitude et la précision. Globalement, l’utilisation de plus d’exécutions semble supérieure, mais l’approche par atlas basé des vérités terrain donne tout de même des résultats similaires à moindre coût.

#### 4.6. RÉSULTATS

# runs	Sensitivity		Specificity		Accuracy		Precision	
	MED	IQR	MED	IQR	MED	IQR	MED	IQR
1	0.91	0.09	0.90	0.04	0.90	0.03	0.65	0.09
2	0.95	0.07	0.88	0.05	0.89	0.03	0.59	0.08
4	0.94	0.08	0.89	0.05	0.89	0.03	0.62	0.10
8	0.94	0.08	0.90	0.04	0.90	0.02	0.65	0.09
16	0.94	0.07	0.90	0.03	0.91	0.02	0.67	0.07
32	0.93	0.07	0.92	0.03	0.92	0.02	0.69	0.09
64	0.93	0.06	0.92	0.03	0.92	0.02	0.70	0.07

tableau 4.6 – Mesure de classification binaire pour la représentation voxélique. Globalement, comme ci-haut, le nombre d'exécutions a une influence positive sur le chevauchement avec la vérité terrain. Par contre, l'approche par atlas basée sur les vérités terrain obtient, à nombre d'exécutions égal, de moins bons résultats que l'approche de pige aléatoire. En effet, cette approche semble moins précise, ce qui veut dire que la sensibilité est augmentée au prix de la spécificité.

### 4.7 Utilisations externes

Bien qu'il ne s'agisse pas d'une quantification de l'algorithme, les qualités qualitatives (respect de l'anatomie, facilité d'utilisation et rapidité d'exécution) ont convaincu de nombreux collaborateurs et mené à l'utilisation de l'algorithme dans le cadre de plusieurs projets.

**MS13 :** Utilisation de l'algorithme pour un projet lié à la sclérose en plaques au Centre Hospitalier Universitaire de Sherbrooke (CHUS). Des modèles personnalisés ont été développés pour améliorer les résultats liés à la tractométrie et minimiser le nombre de tests statistiques.

**HOG :** Utilisation de l'algorithme pour un projet lié aux commotions cérébrales à l'Université de Sherbrooke pour l'équipe de Football du Vert&Or. Des modèles personnalisés ont été développés pour améliorer les résultats liés à la tractométrie et minimiser le nombre de tests statistiques.

**Berkeley :** Utilisation de l'algorithme pour un projet lié aux accidents vasculaires cérébraux à l'Université de Berkeley en Californie. Des modèles personnalisés ont été développés pour satisfaire les besoins des collaborateurs, le faisceau arqué (AF) était sous-divisé en 5, en plus de segmenter une dizaine d'autres faisceaux d'association. Un atlas personnalisé a été utilisé pour ce projet, les segmentations manuelles ont été fournies par le laboratoire.

**Chirurgie :** Utilisation de l'algorithme pour faciliter la visualisation dans le cadre la planification chirurgicale (i.e tumeurs) au Centre Hospitalier Universitaire de Sherbrooke (CHUS). L'algorithme est utilisé pour segmenter automatiquement une vingtaine de faisceaux d'intérêt. Les faisceaux segmentés permettent de rapidement voir si la zone entourant la tumeur a poussé ou «*coupé*» une connexion importante et aide même parfois à estimer le degré d'agressivité de la résection.

## 4.7. UTILISATIONS EXTERNES

**Imeka Solutions :** Dans le cadre d’une déclaration d’invention conjointe avec l’Université de Sherbrooke et Imeka Solution l’algorithme est utilisé pour segmenter automatiquement une vingtaine de faisceaux d’intérêt. Typiquement, ces résultats sont ensuite utilisés comme entrées pour la tractométrie pour quantifier les différences causées par des pathologies, le développement ou autres. L’algorithme a été utilisé par la compagnie sur des projets d’Alzheimer, Parkinson ou sclérose en plaques avec des centaines, voir milliers de participants (essais cliniques).

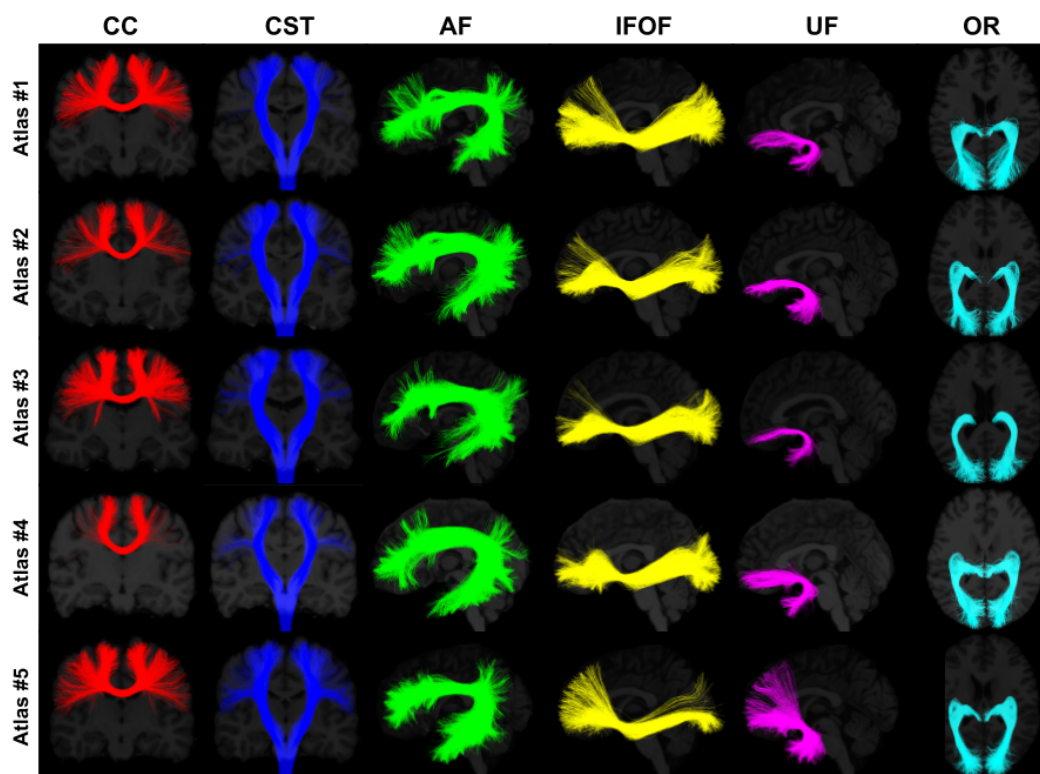


figure 4.8 – Cette grille présente l’atlas utilisé pour de nombreux projets liés au laboratoire. Ces 6 faisceaux (sur 33) chez 5 sujets HCP montrent la variabilité morphologique de l’atlas. La définition anatomique utilisée pour la segmentation est inspirée de [Catani et De Schotten, 2008; Makris et al., 2004; Witelson, 1985], mais adaptée pour la tractographie probabiliste. Comme *RecobundlesX* ne force pas de zones de terminaisons exactes, seulement la morphologie générale est importante. Abréviations : corpus callosum (CC), corticospinal tracts (CST), arcuate fasciculus (AF), inferior fronto-occipital fasciculus (IFOF), uncinate fasciculus (UF), optic radiation (OR).

## 4.7. UTILISATIONS EXTERNES

### 4.7.1 Création d’atlas personnalisés

L’étude du faisceau pyramidal ici présentée avait comme objectif de démontrer l’utilité de l’approche multi-paramètres et multi-atlas de *RecobundlesX*. Les performances exactes de l’algorithme sont difficilement généralisables, mais la tendance entre nombre d’exécutions et amélioration de résultats tiendra. C’est pourquoi un ensemble d’atlas de faisceaux a été créé pour les utilisations ci-dessus (liste non-exhaustive). À partir de tractogrammes probabilistes provenant de 5 sujets HCP, 33 faisceaux ont été segmentés et «*nettoyés*» de manière similaire à la méthodologie présentée à la Section 5.2.5. La Figure 4.8 montre 6 de ces faisceaux chez les 5 sujets HCP.

De façon plus explicite, voici les étapes nécessaires pour créer un ensemble de modèles de faisceaux (atlas) dans un scénario idéal.

#### Définir les faisceaux

En utilisant une méthode classique (e.g segmentation manuelle basée sur des ROIs), isoler les faisceaux d’intérêt de manière à éviter autant que possible les données aberrantes. Ces segmentations devraient être le plus strict possibles :

- Terminaisons à la matière grise
- Cohérence en termes de forme
- 100% des *streamlines* doivent être considérées valides

Cette rigueur est nécessaire, car une seule *streamline* suffit pour que RecobundlesX reconnaisse cette forme dans un tractogramme et que celle-ci fasse désormais partie de la segmentation. Comme RecobundlesX est un algorithme multi-atlas, l’étape présentement décrite peut (doit) être effectuée sur une multitude de jeux de données (plusieurs sujets, plusieurs types d’acquisitions, plusieurs types d’algorithmes).

#### Création des atlas

Le concept *multi-atlas* peut devenir plus complexe lorsque plusieurs sujets de plusieurs sources ont été utilisés. Dans le but de simplifier l’explication, nous allons utiliser un exemple concret. Un projet ayant utilisé 10 sujets pour chacune de 5 bases de données différentes (e.g HCP, UKBioBank, etc.) pourrait vouloir créer 5 atlas



## 4.7. UTILISATIONS EXTERNES

indépendants où chacun sera la moyenne de 10 sujets. À l'étape précédente, nous avons décrit l'importance de la définition initiale de ces faisceaux, assumons que cette étape s'est bien déroulée pour les 50 (5 x 10) sujets.

L'étape suivante consiste à faire un recalage non-linéaire de ces 50 sujets vers un espace commun (e.g mni152) à l'aide de leur image structurelle (i.e T1) et d'ensuite appliquer cette transformation aux faisceaux. Une fois tous dans le même espace, les faisceaux peuvent être fusionnés (concaténés) par base de données ce qui résultera en plusieurs atlas (5) qui représenteront en fait une moyenne de leurs sujets (10).

### Nettoyage des atlas

Une fois les faisceaux fusionnés dans un espace commun, les étapes de nettoyage seront simplifiées et accélérées. Pour chaque faisceau de chaque atlas, il est important de garantir l'absence de *streamlines* indésirables. Il est recommandé d'utiliser une approche par *clusters*, en effet en regardant individuellement les *clusters* il est possible de facilement identifier ceux ayant une forme aberrante :

- Boucle et/ou changement de direction abrupte
- *Streamline* incomplète ou trop courte

Le but de cette opération est de nettoyer de manière stricte, comme à la première étape, mais à l'échelle d'une population. Il est crucial de garantir une forme cohérente sans *streamlines* indésirables.

### Symétrisation des atlas

L'asymétrie des faisceaux de la matière blanche est bien connue, par contre il n'est pas souhaitable que les atlas induisent une asymétrie qui ne provient pas réellement du tractogramme à segmenter. C'est pourquoi il est fortement conseillé de symétriser les atlas utilisés par RecobundlesX. Cette opération relativement simple consiste à créer un effet miroir sur les faisceaux (et l'image de référence en espace commun), ré-aligner l'image de référence (transformation affine) et appliquer cette légère transformation aux faisceaux. L'opération de recalage ne sert qu'à remettre en place les faisceaux si l'image de référence n'était pas parfaitement centrée sur son axe central.

Le résultat de cette opération est plusieurs atlas avec une symétrie gauche-droit

## 4.7. UTILISATIONS EXTERNES

garantie. Donc si une asymétrie est observée après la segmentation automatique de RecobundlesX, celle-ci viendra du tractogramme lui-même.

### Lissage des atlas

Certaines manipulations mineures peuvent être nécessaires. Par exemple dans le cas de données de tractographie probabilistes, un léger effet de zigzag est visible. Ces zigzags dus à la nature même de l'algorithme de tractographie ne transportent pas d'information utile pour RecobundlesX et peuvent donc être enlevés. Une simple opération de lissage le long de la streamline (e.g courbe spline) est suffisante.

De plus, il pourrait être souhaitable de randomiser l'ordre des *streamlines* pour contrer un biais possible de l'opération de concaténation ou de symétrisation. Comme RecobundlesX repose sur l'algorithme Quickbundles [Garyfallidis et al., 2012], l'ordre des *streamlines* pourrait avoir une influence.

### Réduction des atlas

Il est souhaitable de minimiser la taille des faisceaux des atlas pour assurer le bon fonctionnement de RecobundlesX. En effet, l'opération de concaténation des sujets (pour obtenir une représentation d'une population) a généré des faisceaux contenant beaucoup de *streamlines*. Par contre, beaucoup de ces *streamlines* ne transportent pas d'information utile, car elles sont redondantes. Deux *streamlines* quasi-identiques ne sont pas mieux qu'une, et il est donc souhaitable de retirer les *streamlines* similaires.

Considérant les limitations algorithmiques, les faisceaux utilisés devraient posséder moins de 5000 *streamlines* tout en conservant autant que possible leur forme initiale. L'élimination de *streamlines* similaires se fait en utilisant la MDF [Garyfallidis et al., 2010]. Deux *streamlines* ayant une distance inférieure à 2mm-3mm peuvent être considérées équivalente et donc une d'entre elles peut être retirée.

### Évaluation finale des atlas

Ces étapes sont celles recommandées pour la création d'atlas personnalisés, l'exemple précédent va mener à la création de 5 atlas chacun constitué de faisceaux moyens issus

## 4.8. DISCUSSIONS

de 10 individus. L'objectif global est toujours d'obtenir une représentation parfaite des faisceaux couvrant la variabilité possible d'une population.

Il est important de mentionner que toute étape supplémentaire pour atteindre cet objectif est justifiable, la création d'atlas pour RecobundlesX n'est pas une science exacte. Si un projet concerne une population pédiatrique ou gériatrique, il est probablement souhaitable de créer des atlas issus de cette population. Mais cela n'est pas nécessaire s'il peut être garanti que ces populations ne présentent pas de différence de forme avec une population saine.

## 4.8 Discussions

### 4.8.1 Influence du nombre d'exécutions

Il est intéressant de noter que le nombre d'exécutions a une influence positive pour toutes les mesures d'évaluation. Non seulement les résultats médians sont supérieurs, mais la variabilité de ceux-ci décroît. Dans un scénario typique, cela veut dire que le niveau de confiance dans les résultats peut être supérieur, car les résultats ont moins de chances d'être une donnée aberrante piquée au hasard dans l'ensemble des solutions possibles.

Sans nécessiter une nouvelle approche plus complexe, le simple fait de mieux explorer l'espace des résultats possibles est suffisant pour améliorer et stabiliser le processus de segmentation de faisceaux. Dans une situation réelle où cet algorithme (RecobundlesX) est utilisé, il est recommandé d'utiliser entre 32 et 64 exécutions total (e.g 8 atlas et 8 jeux de paramètres) pour surmonter les limitations de Recobundles tout en conservant un temps d'exécution raisonnable.

### Multi-atlas v. multi paramètres

Ces deux facteurs ont été présentés comme un tout, par contre ceux-ci ont été séparés au moment des calculs d'analyse. Lors de l'expérimentation basée sur toutes les segmentations manuelles au niveau de qualité variable, le fait d'utiliser un maximum d'atlas semble avoir la plus grande influence. Par contre, lors de l'expérimentation où les vérités terrain étaient utilisées, et donc leur niveau de qualité très stable, il

## 4.8. DISCUSSIONS

semble que le nombre de jeux de paramètres avait une plus grande influence. Il est aussi important de considérer que pour cette expérimentation le nombre d'atlas était limité à 4 dû au fait que les données venaient du projet Tractostorm.

Dans un contexte où la qualité des entrées peut être garantie par un choix manuel *a posteriori*, il est recommandé de retirer les modèles considérés aberrants si présents. Cela peut en effet éviter des exécutions qui produiront des résultats qui seraient potentiellement inutiles lors de la fusion d'étiquettes.

### Fusion d'étiquettes

Il est important de considérer le fait que pour certaines situations la sensibilité est plus importante que la spécificité. Plutôt que de modifier l'ensemble de paramètres d'entrée ou de modifier les atlas, il est possible de simplement modifier la fusion d'étiquettes. En effet, pour limiter la quantité de calcul cet hyper-paramètre n'a pas été présenté, mais plutôt qu'un vote majoritaire (50%), il serait possible d'utiliser d'autres valeurs. Un vote à 25% sera plutôt libéral, la sensibilité sera plus élevée au prix de la spécificité. Un vote à 75% sera quant à lui plus conservateur, la spécificité sera plus élevée au prix de la sensibilité. Cet hyper-paramètre permet de rapidement modifier la «taille» des résultats, étant donné les sorties de l'algorithme il peut même être modifié *a posteriori*.

### Temps d'exécution

Augmenter le nombre d'exécutions vient évidemment au prix du temps d'exécution. L'implémentation de l'algorithme multi-atlas et multi-paramètres comme une méthode raffinée permet de réduire le temps d'exécution. Les opérations liées au chargement, sauvegarde et les opérations de pré-traitement peuvent être effectuées qu'une seule fois et donc réutilisées. De plus, dans la présente expérimentation seulement deux faisceaux ont été segmentés. Dans une situation où plusieurs faisceaux seraient à identifier, l'avantage d'utiliser un algorithme où le traitement est fait en mémoire et les étapes en commun ré-utilisées est clair.

Les multiples exécutions ont été parallélisées, seule l'étape de fusion d'étiquettes nécessite tous les résultats pour la sauvegarde finale. Cet avantage permet de grande-

## 4.8. DISCUSSIONS

ment réduire le temps d'exécution et est particulièrement intéressant lorsque plusieurs faisceaux doivent être segmentés.

### 4.8.2 Comparaison à l'approche manuelle

Il est important de noter que la comparaison de l'algorithme *RecobundlesX* à la vérité terrain doit s'effectuer en gardant en tête que l'expertise humaine n'est pas parfaite. En effet, les résultats du projet Tractostorm montrent que les segmentations manuelles (experts ou non) possèdent une variabilité non négligeable. Tel que montré dans la Section 4.6 la médiane et l'écart interquartile sont à considérer. Les résultats de *RecobundlesX* sont similaires, voire supérieurs, au groupe non-experts. Pour ce qui est du groupe expert, l'expertise humaine semble plus précise (au détriment de la sensibilité).

Cependant, une question plus incertaine est de savoir si l'algorithme *RecobundlesX* serait toujours comparable à l'expertise humaine. Dans la publication originale de l'algorithme *Recobundles*, la vérité terrain était le résultat de segmentation ne prenant pas en compte la variabilité intra et inter évaluateur. Un tel scénario ne rend pas justice à l'algorithme, mais une investigation telle que présentée dans le projet Tractostorm est très coûteuse. La généralisation des observations ici présentées n'est donc pas triviale.

### Travaux futurs

Une limitation majeure de cette analyse est évidemment l'utilisation d'un seul faisceau, mais la création d'un jeu de données adapté et la complexité de l'analyse antérieure de Tractostorm font que cette première étape est une contribution importante. Par contre, il est évident que des travaux seraient nécessaires pour évaluer au moins un faisceau d'association et commissural. Pour effectuer une telle analyse, un projet similaire à Tractostorm devrait être entrepris pour avoir non seulement un jeu de données adéquat, mais aussi pour avoir un point de comparaison à l'expertise humaine.

Dans un tel scénario, il est fort probable que la tendance de l'amélioration des résultats en fonction du nombre d'exécutions tienne toujours. Les résultats issus de

#### 4.9. CONCLUSION

la littérature sur la segmentation d’images multi-atlas montrent qu’une meilleure recherche de l’espace des possibilités produira toujours des résultats plus robustes, plus stables et généralement de meilleure qualité [Iglesias et Sabuncu, 2015; Lampert et al., 2016; Pipitone et al., 2014]. L’arrivée d’algorithmes de segmentation utilisant la puissance de l’apprentissage machine va possiblement grandement améliorer le domaine de la segmentation de faisceaux. Bien que la représentation par *streamlines* soit encore difficile à gérer, certains travaux préliminaires montrent que la capacité de ces algorithmes à reconnaître des faisceaux est très prometteuse. De plus, cela réduirait la dépendance aux atlas (difficile à construire) tout en accélérant les opérations de segmentations automatiques.

## 4.9 Conclusion

Les résultats présentés démontrent que l’utilisation d’une approche d’exécutions multiple permet l’amélioration des résultats et la réduction de la variabilité de l’algorithme *Recobundles*.

““You miss 100% of the shot you don’t  
take.”  
- Wayne Gretzky”

---

— Michael Scott, The Office.

## Chapitre 5

# Tractographie spécifique aux faisceaux avec *a priori* anatomique et d'orientation

### Résumé

Les faisceaux de la matière blanche varient en termes de forme, taille, longueur et complexité, ce qui fait que les reconstructions par tractographies de certains faisceaux sont plus difficiles que d'autres. Au final, la reconstruction de faisceau souffre parfois d'une couverture spatiale inadéquate. Pour remplir l'espace de la matière blanche autant que possible, des millions de *streamlines* doivent être générées et ensuite filtrées pour contourner cette limitation. Par contre, certains biais sont par le fait même introduits, des biais tels que l'abondance de faux positifs et la surreprésentation de faisceaux faciles à reconstruire et la sous-représentation de faisceaux difficiles à reconstruire. Pour surmonter ces défis, nous avons développé un algorithme de tractographie spécifique par faisceau. Cet algorithme incorpore des *a priori* de positions et d'orientations dans le processus de génération de *streamlines* pour augmenter la reproductibilité, la sensibilité, la spécificité et l'efficacité des reconstructions de faisceaux d'intérêt. La tractographie spécifique par faisceaux performe mieux que les méthodes classiques telles que les algorithmes déterministes, probabilistes et globaux.



L'augmentation de *streamlines* anatomiquement plausibles, avec plus de couverture spatiale, aide à obtenir une représentation plus exacte de la morphologie globale des faisceaux, ce qui améliore et robustifie les études d'imagerie basées sur la connectivité.

## Contributions de la publication

- Introduction du concept de tractographie (classique) par faisceaux, directement dans l'espace des *streamlines*.
- Utilisation de l'objet *TODI* (Tract orientation density imaging) appliqué par faisceaux comme *a priori*.
- Cadre pour la génération de patron de faisceaux.
- Formalisation du biais de densité et proposition de solutions pour la génération de patrons.

## Contributions des auteurs

- Développement de la méthode (Francois Rheault et Etienne St-Onge)
- Analyse, écriture et création des tables et figures (Francois Rheault)
- Support pour l'interprétation anatomique des résultats (Jasmeen Sidhu et Laurent Petit)
- Révision du manuscrit (Klaus Maier-Hein, Nathalie Tzourio-Mazoyer)
- Supervision du projet (Laurent Petit, Maxime Descoteaux)

## Commentaires

L'idée derrière ce projet a émergé d'un besoin bien particulier de notre collaborateur, M. Laurent Petit à l'Université de Bordeaux. Son objectif était de construire un atlas probabiliste du faisceau pyramidal chez plus de 400 sujets, mais les résultats n'étaient pas optimaux, loin de la réalité anatomique connue et surtout coûteuse en temps de calcul. Etienne St-Onge et moi avons donc concocté un algorithme dédié à améliorer le modèle local (fODF) à partir d'*a priori* d'orientation automatiquement générée provenant d'un modèle de faisceaux. M. Laurent Petit a été rapidement impressionné par les résultats et notre méthode utilisée dans la publication qui en a

découlé [Chenot et al., 2019]. De notre côté l’algorithme a été soumis à la conférence MICCAI, mais suite à un oubli de ma part nos noms ont été laissés sur la soumission et l’article refusé. Grâce à Chantal Tax, nos travaux ont pu survivre dans un chapitre du Computational dMRI de MICCAI. Ensuite, nous avons développé notre méthode, élargi notre analyse et nous avons dû améliorer notre contexte/introduction suite à la découverte d’une publication similaire d’un groupe allemand (dont le directeur a fini par joindre notre publication). La publication s’est donc retrouvée dans le journal Neuroimage (soumise le 27 août 2018, acceptée le 13 novembre) et notre méthode a déjà été utilisée par de nombreux collaborateurs.

# Bundle-Specific Tractography with incorporated anatomical and orientational priors

Francois Rheault<sup>1</sup>, Etienne St-Onge<sup>1</sup>, Jasmeen Sidhu<sup>1</sup>, Klaus Maier-Hein<sup>3</sup>, Nathalie Tzourio-Mazoyer<sup>2</sup>, Laurent Petit<sup>2</sup>, Maxime Descoteaux<sup>1</sup>

<sup>1</sup>Sherbrooke Connectivity Imaging Laboratory (SCIL), Université de Sherbrooke, Canada

<sup>2</sup>Groupe d'Imagerie Neurofonctionnelle, IMN, UMR5293, Université de Bordeaux, France

<sup>3</sup>Division of Medical Image Computing, German Cancer Research Center (DKFZ), Heidelberg, Germany

**Keywords:** Diffusion MRI, White Matter, Tractography, Tractometry, Connectivity

## Abstract

Anatomical white matter bundles vary in shape, size, length, and complexity, making diffusion MRI tractography reconstruction of some bundles more difficult than others. As a result, bundles reconstruction often suffers from a poor spatial extent recovery. To fill-up the white matter volume as much and as best as possible, millions of streamlines can be generated and filtering techniques applied to address this issue. However, well-known problems and biases are introduced such as the creation of a large number of false positives and over-representation of easy-to-track parts of bundles and under-representation of hard-to-track. To address these challenges, we developed a Bundle-Specific Tractography (BST) algorithm. It incorporates anatomical and orientational prior knowledge during the process of streamline tracing to increase reproducibility, sensitivity, specificity and efficiency when reconstructing certain bundles of interest. BST outperforms classical deterministic, probabilistic, and global tractography methods. The increase in anatomically plausible streamlines, with larger spatial coverage, helps to accurately represent the full shape of bundles, which could greatly enhance and robustify tract-based and connectivity-based neuroimaging studies.

## 5.1 Introduction

Tractography algorithms are currently challenged by several important open issues and biases that need to be addressed to improve brain connectivity mapping [Jbabdi et al., 2011; Jones, 2010; Maier-Hein et al., 2017]. In Maier-Hein et al. [2017], it was recently shown that whole-brain tractography can recover most existing true positive bundles but that these are poorly recovered in terms of spatial extent. In the simulations of 25 human-like synthetic white matter (WM) bundles, an average spatial extent recovery of only 30% is reported [Maier-Hein et al., 2017, Fig. 3]. We propose to term this tractography issue as the “*poor-spatial-extent*” challenge. Techniques that have better spatial extent recovery always generate more invalid streamlines, and thus fall into the usual *sensitivity-specificity curse*. In the current context, the notion of invalid streamlines refers to anatomically improbable streamlines, meaning their path is not reflective of known WM pathways. The quality of WM bundle reconstruction using tractography is influenced by many variables : i) global geometry of the bundle like curvature, thickness, length, ii) local crossing regions it traverses, position of the bundle with respect to other bundles, and iii) extent and shape of cortical endpoints. As a result, some WM bundles are harder-to-track than others, which leads to the poor-spatial-extent challenge.

Why are some bundles more difficult to reconstruct than others? Complex local fiber regions, even if well recovered locally in terms of fiber orientation distributions, can lead to invalid tractography streamlines that respect the underlying local diffusion signal but make clear invalid pathways at the global brain level [Maier-Hein et al., 2017]. Despite the fact that almost all WM regions contain crossing fiber populations [Jeurissen et al., 2013], some are well-know pitfalls for tractography methods and could be tackled with better prior information. As an illustration, the *centrum semiovale*, as seen in Figure 5.1 and Figure 5.2, was and still is a major difficulty for tractography algorithms. Three major WM pathways : i) the pyramidal tract, ii) the mid-body of the corpus callosum (anterior and posterior), and iii) the arching streamlines connecting the frontal to the temporal lobes, are crossing almost orthogonally in the most ventral part. At such an intersection, even with multiple fiber techniques, tractography algorithms most often follow the easiest path available. Bundles can

## 5.1. INTRODUCTION

seem to cover a large spatial extent and have an appropriate amount of fanning when millions of streamlines are used. However, the density is far from uniform within a bundle, even if it cannot be perceived visually, as seen in Figure 5.2. The amount of crossing regions, overall length and curvature of the bundle will influence the probability of a streamlines to reach its expected destination. Moreover, it is common to observe streamlines abruptly changing direction to either switch to a different bundle or follow an erroneous path, but end up in the right location. Crossing can become even harder to distinguish as the main direction of the fiber populations kiss or come into a bottleneck [Maier-Hein et al., 2017, Fig.7].

A partially reconstructed WM bundle could lead to erroneous interpretations and conclusions. At the tract-based analysis level [Cousineau et al., 2017; Dayan et al., 2016; Yeatman et al., 2012, 2018], the spatial extent of bundles (in  $\text{cm}^3$ ) is a measurement often used to investigate morphological group differences, such as atrophy or asymmetry [Catani et al., 2007; Chenot et al., 2018; Song et al., 2014]. At the connectome-level, the number of streamlines in WM bundles connecting all possible pairs of cortical regions is usually used [Sotiropoulos et Zalesky, 2019]. However, because of a combination of local and global differences in brain geometry and WM configurations, certain subjects may have easier-to-track bundles than other subjects, which results in WM bundles that fill-up in spatial extent before others and thus have artificially larger number of streamlines. Two strategies exist to attempt overcoming these issues : 1) a whole brain tractography with millions of streamlines followed by manual or automatic filtering techniques, 2) a tailored ROI-based tractography.

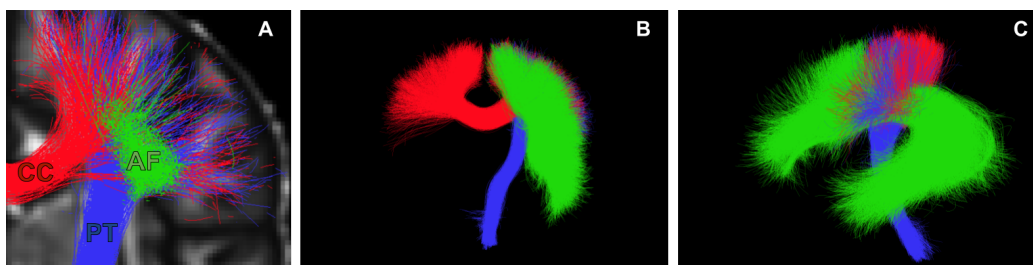


figure 5.1 – Crossing region of the Corpus Callosum (red), the Arcuate Fasciculus (green) and the Pyramidal Tract (blue) in the centrum semiovale. A) Coronal cross-section, B) coronal 3D view and C) sagittal 3D view.

## 5.1. INTRODUCTION

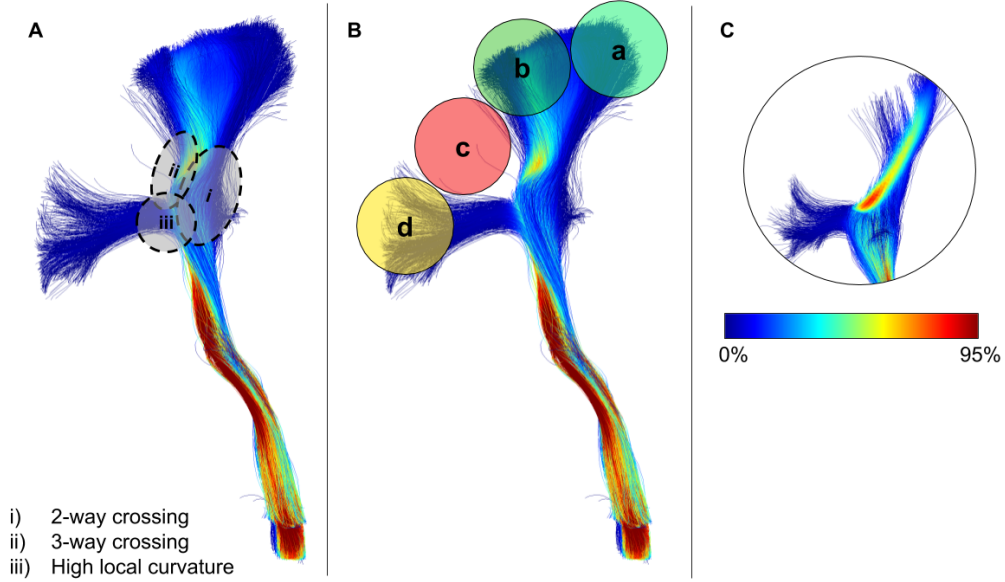


figure 5.2 – Reconstruction of pyramidal tract (PT) streamlines illustrating some of the challenges of tractography, seen on panel A, i) 2-way crossing ii) 3-way crossing, and iii) high curvature. In panel B, regions *a* and *b* encounter less difficulties in reconstruction due to only being subject to a 2-way crossing limitation, as illustrated in Panel A (i), leading to a higher proportion of streamlines reaching the cortex in these general regions. Conversely, region *c* in Panel B encounters 2 and 3 way crossings (Panel A (i) and (ii)), whereas, region *d* in Panel B encounters 2-way crossing and high-curvature limitations (Panel A (i) and (iii)), leading to a lower proportion of streamlines successfully reaching the cortex. Streamline coloring represents the local streamlines density (at the voxel-level), segments in red shows regions where a large proportion of streamlines pass through a bottleneck as opposed to segments in blue. Panel C illustrates that the majority of streamlines traverse through the core of the PT.

### 5.1.1 Whole-brain tractography

In the first strategy, millions of streamlines are typically reconstructed using a whole-brain deterministic, probabilistic, or global fiber tracking algorithms to fill-up the whole WM volume and be able to capture a large spatial extent of all bundles [Calamante et al., 2015; Jeurissen et al., 2017; Moyer et al., 2018; Smith et al., 2015a; Wassermann et al., 2016]. With more streamlines, a larger extent of bundles is reconstructed (higher sensitivity) but also more false positive bundles created (lower

## 5.1. INTRODUCTION

specificity) [Côté et al., 2013; Maier-Hein et al., 2017], thus leading to the *sensitivity-specificity curse* mentioned before. Generation of millions streamlines is often considered a brute force approach with a heavy computational burden. All *iterations* leading to false positive are considered wasted. Reducing such iterations increases the computational efficiency of a method, meaning that less effort is required while maintaining an acceptable sensitivity-specificity ratio.

Moreover, when millions of streamlines are desired in a whole brain tractogram, millions of seeds and hundreds of millions of 3D points are generated, which can be an important computational and memory limitation [Rheault et al., 2017]. In the case of a specific tract-based application, where only a single to a few WM bundles are desired, this whole brain strategy can be sub-optimal both in terms of bundle reconstruction quality and computational efforts. Increasing the number of streamlines to 10M or 100M would fill-up more of this hard-to-track fanning space, but would potentially never reach full extent coverage. Generating more streamlines will artificially increase the density of the central, easy-to-track part of the bundle and generate a large number of false positive streamlines. Advanced filtering techniques such as SIFT, LIFE, COMMIT [Daducci et al., 2015; Pestilli et al., 2014; Smith et al., 2015b] are able to filter some false positives and reduce the density bias by making sure streamlines explain the desired diffusion signal, but at the cost of removing some hard-to-track valid streamlines as well [Maier-Hein et al., 2017].

### 5.1.2 ROI-based tractography

In the second strategy, a ROI-based tractography strategy uses different masks, filtering ROI pass-ways and varying tractography parameters (step size, curvature, anisotropy thresholds) [Chamberland et al., 2017, 2014] to enhance the quality of results based on neuroanatomical prior knowledge. A ROI-based seeding strategy is an efficient way to quickly target a bundle of interest by removing unnecessary computations needed in whole brain tractography. Such strategy has already been used in multiple research projects to avoid generating unwanted streamlines [Basser et al., 2000; Behrens et al., 2007; Catani et al., 2002; Chamberland et al., 2017; Ghaziri et al., 2015; Mori et van Zijl, 2002; Renauld et al., 2016; Rozanski et al.,

## 5.1. INTRODUCTION

2017], mainly using manual delineations, or ROIs obtained from a segmentation tool such as FreeSurfer [Desikan et al., 2006], or the JHU template [Mori et al., 2005].

Here, ROI-based tractography can be more efficient but not necessarily lead to better spatial extent recovery. The poor-spatial-extent problem is temporarily “hidden” because one can now afford to launch millions of seeds just to reconstruct a single bundle. However, the same issues illustrated in Figure 5.2 remain, i.e. hard-to-track fanning regions are hard to reach and easy-to-track regions are amplified. Also, because of the ROI-based filtering process, all the false positive streamlines generated are thrown out and excluded from the result, but, computational efforts are still deployed to track them. Hence, overall, a ROI-based approach still suffers from poor efficiency, incomplete spatial-coverage and a similar underlying trade-off between sensitivity and specificity.

To address these issues, tractography needs the injection of more information and priors into the streamline tracing process itself. In Maier-Hein et al. [2017], it was clearly shown that following local orientations alone cannot resolve all the potential ambiguity and biases of fiber tracking. Bottlenecks are not tractable using directional information alone.

Several approaches have been proposed to inject more prior knowledge. An automated method called TRACULA [Yendiki et al., 2011] was proposed to include anatomical priors from anatomical labeling. Pathway labels are used to initialize and constrain the probabilistic tractography algorithm. However, this approach uses a conservative reconstruction model (FACT, [Mori et al., 1999]) and strict anatomical definitions as training sets, which limits the spatial extent of bundles. Further, probability maps are computed instead of streamlines potentially limiting tract-based analyses. More recently, Wasserthal et al. [2018a] proposed a method to directly segment the volume occupied by bundle of interests. However this method does not rely on tractography and does not produce streamlines. Features of the raw diffusion signal are learned and regions of interests segmented where the characteristics of the signal match with bundles of interests.



## 5.1. INTRODUCTION

On the other hand, global tractography algorithms [Christiaens et al., 2015; Kreher et al., 2008; Mangin et al., 2013; Neher et al., 2012] can sometimes overcome difficult local crossing regions, but resulting tractograms do not always represent the expected anatomical complexity of fasciculi. Global tracking algorithms are promising but are known to have their share of open challenges such as heavy computational needs, the generation of a large number of broken streamlines, and difficulty to impose anatomical priors [Jeurissen et al., 2017; Maier-Hein et al., 2017].

Other approaches exist to overcome fiber crossing difficulties. Chamberland et al. [2017] proposed a magnetic tracking (MAGNET) tool to manually influence directions in strategic regions, improving the reconstruction of the optic radiations bundle. MAGNET essentially pulls tractography towards the occipital lobe with an orientation prior, allowing fiber tracking to perform a “U-turn”-like reconstruction needed to fully reconstruct Meyer’s loop. Moreover, Dhollander et al. [2014] also suggested a method to sharpen orientations based on streamline distributions, called *Track-Orientations Distributions (TOD)*. TOD can help enhance tractography to cover longer distances along WM structures and address some of the biases of fiber tracking. However, since streamlines density is not uniform across a tractogram and does not represent the underlying anatomy, using such a method iteratively across WM will increase the density bias in easy-to-track regions. More recently, Wasserthal et al. [2018b] proposed an approach using bundle-specific orientation maps learned from multiple datasets to select a single fiber orientation distribution (FOD) peak in each voxel, thus improving reconstruction of bundles by simplifying regions with complex WM configuration. To the best of our knowledge, these techniques are the only existing methods trying to inject orientational prior knowledge into the tractography process. Other approaches have proposed to include an orientation-based regularization term in the local reconstruction of fiber orientations, but to our knowledge, not in the tractography procedure itself [Portegies et al., 2015; Reisert et Kiselev, 2011].

Tractography is therefore currently blind to the anatomy and does not have access to the anatomical knowledge accumulated over the past several hundred years by the neuroanatomy community. This neuroanatomical information, when available, could enhance tractography. For instance, priors could capture the “where” and “how” of

## 5.2. METHODS

bundles. That is where should the origin and target region of bundles be, and how difficult crossing, fanning, bending regions should be traversed. Our novel Bundle-Specific Tractography (BST) algorithm is designed to directly incorporate these anatomical and orientational priors in the tractography process itself. The main contribution is a novel methodology enhancing local fiber orientation distribution based on WM bundle priors coming from templates of streamlines. BST is fully automatic and yields better spatial coverage, increases quality in the fanning extents and produces more plausible shape reconstructions of bundles in the centrum semiovale.

## 5.2 Methods

### Bundle-Specific Tractography

Our novel bundle-specific tractography (BST) approach is composed of three steps. For each bundle of interest to be reconstructed, one needs to :

1. Build a ***template of streamlines*** that represents the shape and position of each bundle, covering as much geometric variability as possible (see first row of Figure 5.3 and Figure 5.4).
2. Build the ***anatomical priors*** by incorporating dilated versions of endpoints and spatial coverage maps defined from the template of streamlines in 1) (see second row of Figure 5.3).
3. Build the ***orientation priors*** by generating a field of enhanced fiber orientation distributions (E-FOD) from the track-orientation distribution of the template of streamlines in 1) (see Figure 5.4).

Our BST approach can be applied to any bundle of interest with the previous 3 steps. We now detail each of these steps before describing the datasets and how it was applied to bundles intersecting in the centrum semiovale seen in Figure 5.1.

## 5.2. METHODS

### 5.2.1 Building the templates of streamlines

This initial step is crucial since the template of streamlines affects all additional orientational and anatomical priors used by the BST algorithm. It involves four technical steps (see Figure A.1) :

1. Run a whole-brain or ROI-based tractography technique on the selected subjects.
2. Segment the bundles of interest.
3. Merge all streamlines of each bundle and each subject in a common space.
4. Adjust the density bias of streamlines distribution.

**i) Initial tractography** Our approach works in combination with any existing whole-brain streamline-based or iterative tractography algorithm [Jeurissen et al., 2017]. BST will enhance the initial tractography algorithm and retain the properties

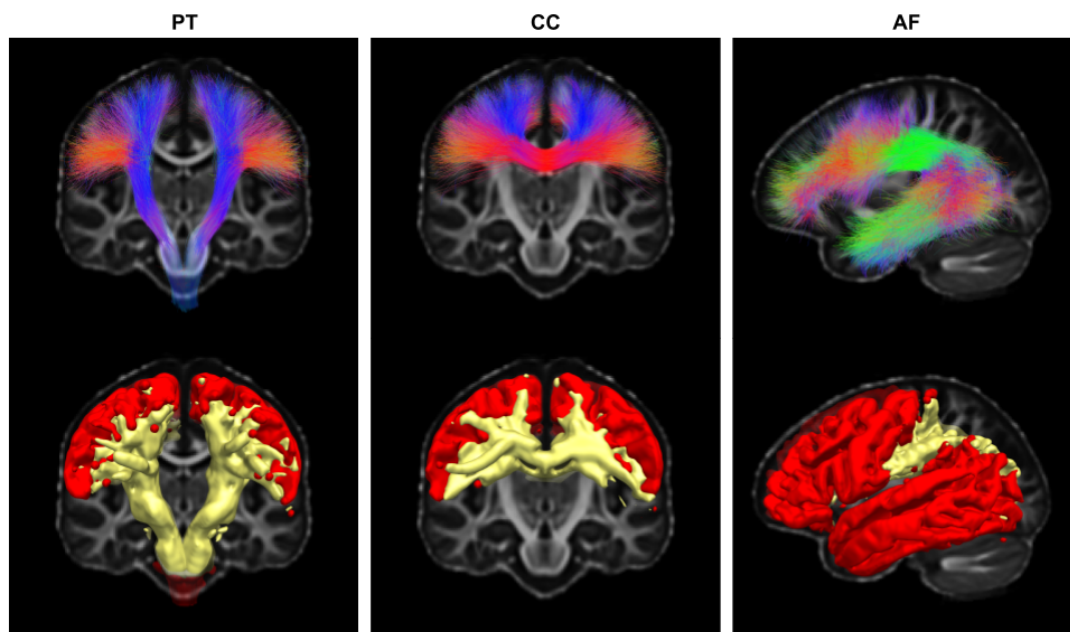


figure 5.3 – The first row shows streamlines from the template on top of the FA template. In the second row, tracking masks (yellow) and seeding endpoints masks (red), automatically extracted from the template of streamlines of the first row.

## 5.2. METHODS

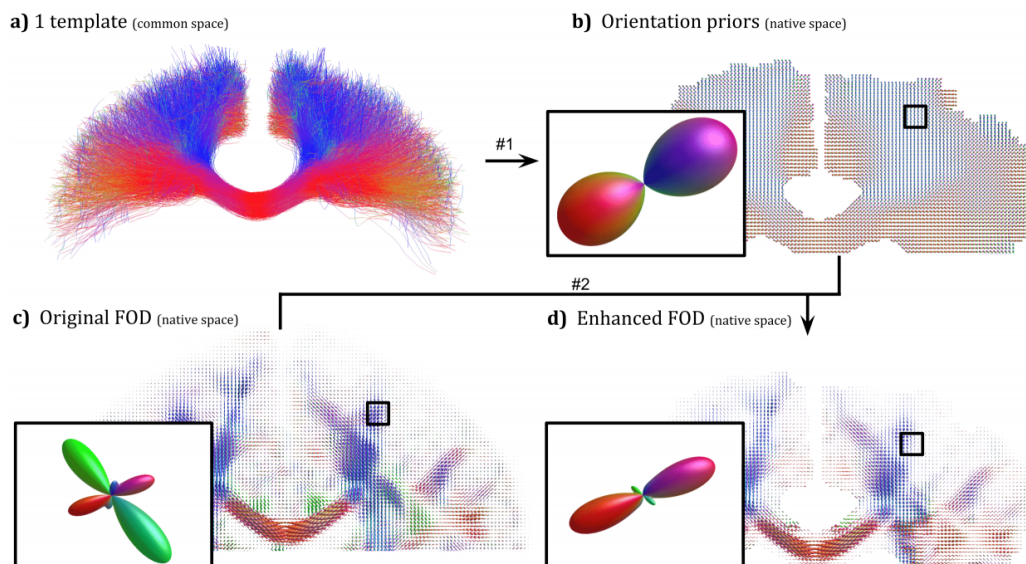


figure 5.4 – a) Template in a standard space, b) orientation priors in native space, c) initial FOD from CSD, d) pre-computed enhanced FOD.

of the chosen initial algorithm. Global tracking and graph-based/geodesic techniques could also be casted in a BST setting but this is outside the scope of the current paper and will be discussed later.

**ii) Segmentation of bundles** Bundles of interest needed for a particular study are usually well-defined in terms of anatomical regions they traverse or connect to [Catani et De Schotten, 2008; Catani et al., 2002]. Moreover, the shape of the bundle is generally agreed upon by various experts [Catani et De Schotten, 2008; Hau et al., 2016] or have often been established from *ex vivo* Klingler blunt dissections [Benedictis et al., 2016; Fernández-Miranda et al., 2015; Hau et al., 2017]. Hence, this segmentation step of bundles is either done manually or with a semi-automatic filtering or clustering approach. Of course, if the number of subjects is large, a manual approach can be tedious and prone to human error. Using ROIs defined by in WM and GM atlases to specify a sequence of rules describing a pathway can be a good alternative [Wassermann et al., 2016]. Another segmentation method is to select or discard streamlines based on shape of cluster using models or anatomical *a priori* [Chekir et al., 2014; Garyfallidis et al., 2018; Guevara et al., 2011; O'donnell et al.,

## 5.2. METHODS

2013; Voineskos et al., 2009; Zhang et al., 2018a].

**iii) Bring streamlines to a common space** A common space needs to be chosen. The usual MNI space is a possibility or a tailored template space defined for a specific group of subjects. Each subject is nonlinearly warped to that common space and the deformation field used to move each bundle of interest from each subject into the common space. The deformation involves transforming each point (vertex) of each streamline, using *tract-math* [Wassermann et al., 2016] or a in-house tool. Once aligned to a common space, bundles of interest are concatenated together into single track file, as seen in Figure 5.3. The intention of these templates is to represent the general shape and position of bundles.

**iv) Correct for the density bias** An individual subject tractogram generally has low density in crossing regions, while extremely dense in regions of one fiber population (easy-to-track parts). The overall shape of a bundle influences this density, i.e. straight bundles will be denser than curved bundles. Since the concept of tract density (streamlines count) in tractography can be a major pitfall that does not accurately represent the underlying WM structure [Daducci et al., 2016; Jbabdi et Johansen-Berg, 2011; Jones, 2010], this non-uniform density bias needs to be removed from the template of streamlines as only the orientation is of interest for our priors. A smart streamline subsampling method based on hierarchical clustering algorithm and minimal direct flip (MDF) distance [Garyfallidis et al., 2012, 2016] is used to overcome this issue. A streamline within 2mm of its neighboring streamlines do not carry new information and is removed, without altering the general shape of the template.

As mentioned before, the template of streamlines essentially captures the a priori knowledge of “*where*” and “*how*”, i.e. where is the bundle originating from, where is going to, and how is it supposed to traverse complex crossing/kissing/branching/fanning regions.

## 5.2. METHODS

### 5.2.2 Anatomical priors : Tractography map creation

The streamlines in the template can now be displaced using the same point-wise method from the template creation (using the inverse deformation field). Once the template of streamlines is in the subject space, a new tracking mask is automatically generated from the voxels intersected by the template streamlines. The streamlines endpoints can also be used to generate a ROI seeding mask. To reduce gyral bias effects and to cover the potential variability across subjects, the seeding and tracking regions are all dilated by 5mm, as seen in Figure 5.3. These masks are used to generate probabilistic exclusion and inclusion maps as well as an interface GM-WM map needed for anatomically-constrained tractography [Girard et al., 2014; Smith et al., 2012]. To prevent streamlines from being generated outside of the white matter, all tracking masks were constrained by the subject’s white matter mask. This ensures that our method does not enforce its priors over a subject tissue segmentation, and only the general position within the white matter.

### 5.2.3 Orientational priors : Enhancing fiber orientations

During tractography, streamlines are generated by taking steps using the fiber orientation distribution (FOD) to choose the next direction [Jeurissen et al., 2017]. The value of the FOD is used to weigh the probability of picking a direction. The proposed BST approach uses orientation priors estimated from the template of streamlines to modify the weighting according to the general, or more global, *a priori* bundle direction. If a streamline is following a specific fiber population, the most well-aligned FOD lobe needs to be prioritized.

To achieve this “prioritization” of FOD lobes, the local orientation histogram from nearby segments of streamlines is computed at every voxel. This method is based on a TOD map estimated from the template of streamlines, as originally introduced [Dhollander et al., 2014]. Improvements related to the FOD field aiming to constraint tractography spatially were proposed in the past [Portegies et al., 2015; Reisert et Kiselev, 2011].

## 5.2. METHODS

Then, at each voxel, a point-wise multiplication of the FOD and TOD orientation priors is performed, followed by a normalization. As a result, when a lobe is well-aligned with the directions of the priors, the values stay the same. This is observed in Figure 5.5, where the top row shows TOD maps associated to the pyramidal tracts, the corpus callosum and the arching streamlines respectively. The bottom row shows the corresponding enhanced FOD (E-FOD) associated with each bundle intersecting through the centrum semiovale. Therefore, when a streamline reaches this crossing region, the probability of choosing an appropriate direction is increased. The orientational prior will amplify the desired direction present in the TOD map, and thus improve the directionality of streamlines at the crossing. This concept of using *a priori* information as well as the observed signal to increase anatomical validity of the decision is borrowed from Bayesian statistics, which is common in neuroimaging [Friston et al., 2002a,b; Woolrich et al., 2009].

If the template of streamlines have crossing or complex fanning, the TOD map will correctly capture this and an appropriate weighting will be achieved. This can be seen in the zoomed picture of Figure 5.4. On the contrary, lobes that are perpendicular decrease in value as seen in Figure 5.6. It is important to note that other lobes are not removed, they are simply decreased in amplitude as this information was still present in the diffusion data. As the aim is not to enforce our orientation priors, but simply re-weighting the FOD amplitude, these two scenarios show the impacts in an extreme case and a typical situation.

To ensure that non-existent or erroneous information is not created, a minimal lobe-value threshold is set on the sphere for the orientational priors. This is to ensure that if there is no well-aligned lobe with an amplitude high enough compared to the others, the operation will not increase the values in that direction. A similar threshold was used in Raffelt et al. [2017] to exclude orientations that do not likely represent a fiber population, and used to perform peak extraction on FOD, also to exclude unlikely fiber populations [Chamberland et al., 2014; Dell’Acqua et al., 2013; Tournier et al., 2012]. In our case, we set a relative amplitude threshold to 0.2 (or 20%).



## 5.2. METHODS

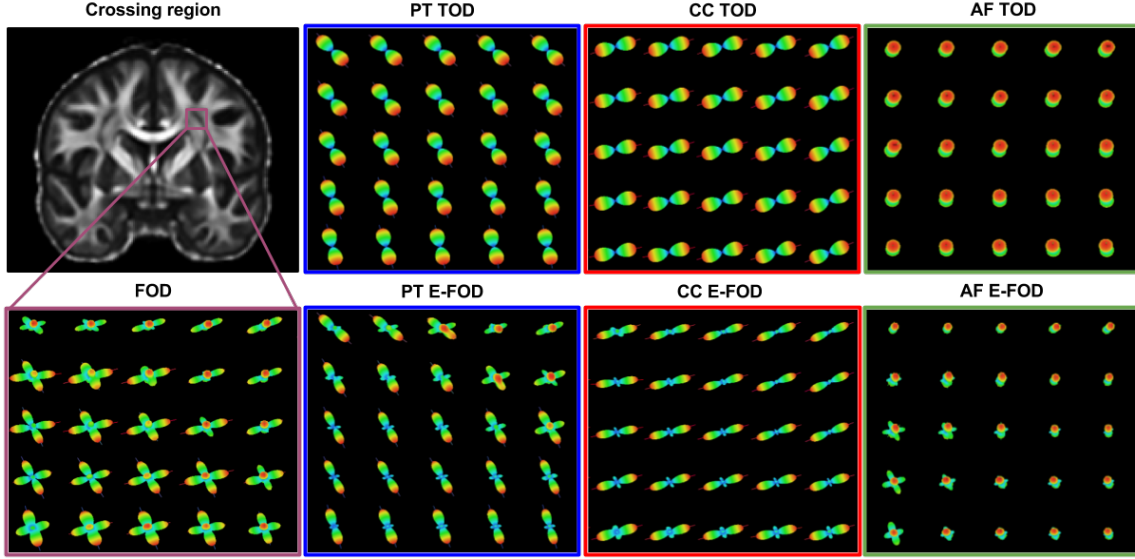


figure 5.5 – Orientation distribution of the crossing section presented in Figure 5.1. The first row illustrates the track orientation distribution (TOD) maps generated from the different template (Pyramidal tract, corpus callosum, Arcuate Fasciculus). In the second row, the first FOD vignette presents the original fiber orientation distribution (FOD) in the crossing region followed by enhanced FOD (E-FOD), a combination of the associated template TOD (from the first row) and the original FOD.

### 5.2.4 DWI Datasets

For this work the templates and experiments were performed with the BIL&GIN database [Mazoyer et al., 2016]. The datasets are composed of T1-weighted and diffusion-weighted images from 39 healthy subjects. The dMRI acquisition consists of 21 gradient directions on half a sphere, the same direction were acquired on the other half of the sphere for averaging, on a single shell ( $b = 1000 \text{ s/mm}^2$ ) with a  $2 \text{ mm}$  isotropic resolution. This scheme was acquired twice for averaging to increase SNR. At each voxel, the FOD was computed using constrained spherical deconvolution (CSD) with spherical harmonics of order 6 Descoteaux et al. [2009]; Tournier et al. [2008].



## 5.2. METHODS

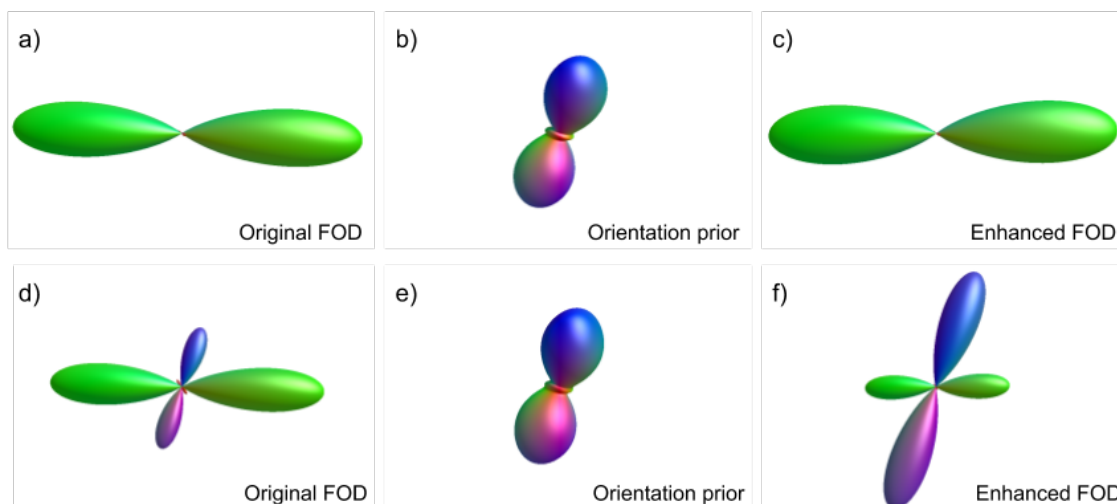


figure 5.6 – The first row represents a voxel where the FOD (a) does not have any lobe aligned with the orientation prior (b), which results in an enhanced-FOD (c) almost identical to the original. The second row represents a voxel where a FOD (d) with a lobe of sufficient size (above the minimum relative amplitude threshold) globally aligned with the orientation prior (e), which results in an enhanced-FOD (f) with a increase amplitude for that lobe.

### 5.2.5 Template creation

For each subject, a whole brain tractogram was reconstructed using the anatomically constrained probabilistic particle filtering tractography (PFT), with recommended default parameters [Girard et al., 2014]. Each tractogram was then segmented using an automatic ROI-based method to extract five bundles of interest, these five bundles will simply be referred to as the left/right pyramidal tract (PT), the corpus callosum (CC) and the left & right arcuate fasciculus (AF). The complete anatomical definitions used by the neuroanatomy expert can be found in the [BST, anatomical definition](#) Section. We built our own tailored anatomical template using the fractional anisotropy (FA) maps of the 39 subjects with ANTS Multivariate Template Construction script [Avants et al., 2008]. This group average template was used as a common space for the template of streamlines. Subsequently, the 5 bundles of interest, from each subject, were moved into that common space using the nonlinear deformation field from ANTS. Once BOIs are concatenated across subject, any redundant stream-

## 5.2. METHODS

lines, up to a 2mm MDF distance threshold, are discarded. Moreover, in this work, a leave-one-out approach was used to ensure that the template streamlines generalize to independent datasets. Any experiment using a template performed on a particular subject did not use the subject itself in that template.

### 5.2.6 Experiments, Evaluation, and Statistical analysis

Probabilistic (Prob) and deterministic (Det), with or without PFT [Girard et al., 2014] were compared, as well as global tractography from *MITK Diffusion* [Neher et al., 2012]. GM-WM interface seeding was performed, with 5 seeds per voxel, adapted to each bundle of interest (BOI) seeding region (BOI map). Three main sets of parameters were used to compare between classical tracking algorithm and the proposed BST method. To distinguish the impact of improvements from bundle-specific tracking anatomical prior masks and orientational prior enhanced-FOD, three tractography reconstructions were performed :

1. Original FOD with the original exclude/include maps
2. Original FOD with bundle-specific exclude/include maps
3. Enhanced FOD with bundle-specific exclude/include maps

Item 1 corresponds to classical HARDI tracking [Jeurissen et al., 2017; Tournier et al., 2012]. The last item is the full BST method, while item 2 is a point of comparison to help separate the contributions of the anatomical and orientation priors. Finally, we also run global tractography [Kreher et al., 2008; Neher et al., 2012] to test against the proposed BST methods. Detailed descriptions of the experiments are available in the [BST, anatomical definition](#) Section. These lead to five bundles, two tracking evolution approaches (Prob and Det), two algorithm (PFT/no-PFT), and the three compared method (classical, BOI map and BOI & E-FOD), for a total of 60 tractography files plus one global tractography result per subject. With these 2379 tractograms (61 tractograms times 39 subjects), the automatic bundle segmentation method mentioned earlier ([BST, anatomical definition](#)) was applied to obtain each bundle.

## 5.2. METHODS

Evaluation and statistical analysis were performed using the following measures : bundle volume, weighted-Dice coefficient, average streamlines length, percentage of valid streamlines and computational performance. The volume was obtained by computing a binary mask of every voxel intersected by streamlines, in a 1mm isotropic grid [Rheault et al., 2017]. The weighted-Dice was introduced by Cousineau et al. [2017] to obtain a value representing spatial agreement that takes account streamlines density. Using the Dice coefficient as a spatial agreement metric will be severely penalized for spurious streamlines. The overall shape of bundles of interests were expected to be similar across subjects. The strict anatomical definitions lead to comparable bundles across subjects once registered in a common space, especially for major WM pathways. In addition, an intra-subject approach would not have provided direct information regarding the quality of a bundle reconstruction. Lack of fanning reaching the cortex could be reproducible and the spatial agreement high without showing improvement compared to the expected shape. Since the bundles of interests in this project (AF, CC, PT) are major WM pathways, they are generally anatomically comparable across subjects, i.e. the spatial agreement is expected to be high, without necessarily reaching a perfect score.

The percentage of valid streamlines is the proportion of generated streamlines respecting the anatomical definition of a bundle from the expert (see [BST, anatomical definition](#) Section) and a length threshold compared to the total amount of streamlines that were originally generated. All outputs from tractography have to respect the algorithm constraints (tracking mask, angular threshold, spherical function threshold), however the length threshold was applied only for bundles. segmentation. In the context of our experiments, the notion of valid and invalid streamlines vary with the bundle of interest investigated. A valid streamline for a particular BOI will be considered invalid for all other BOI. For example, a BOI interface of 100 voxels with 5 seeds per voxel would launch 500 streamlines, if 25 streamlines are segmented (considered valid), this would amount to 5% of valid streamlines. Computational performance (efficiency), represents the total number of tractography iterations, i.e. tracking steps, leading to valid streamline points. This is slightly different than the previous percentage of valid streamlines measure, especially with the BOI tracking mask, because most invalid streamlines end up having less than 50 steps while valid

### 5.3. RESULTS

streamlines end up with more 500 steps. This metric is better than streamlines count to represent the improvement in terms of computational performance.

For statistical analysis, all subjects and bundles were analyzed separately. Then, to highlight general trends, results were averaged. The analysis was performed with all streamlines registered in the FA template space. This alignment, allowed to compute overlaps between subjects and to normalize the values obtained for the volume and average length. This procedure ensures that variations in volume or length does not come from disparities in brain size of the subject but rather from the tracking itself.

## 5.3 Results

To simplify reading of tables, graphs and figures, the focus will be on probabilistic tracking with particle filter tractography (PFT), but all evaluation data, graphs and figures are available in the [BST, anatomical definition](#) Section. Observations and trends are similar and as expected. Moreover, MITK Global tractography was also benchmarked on the same measures to ensure an adequate comparison to state-of-the-art tractography algorithms.

Figure [5.7](#) shows segmentations of PT, CC, and AF bundles of interest with a probabilistic PFT algorithm, comparing the use of original FOD and PFT maps to the proposed BST approach using the enhanced FOD with the modified PFT maps. The reconstructed bundles of interest from the classical tracking show some degree of fanning, but the number of streamlines generated was not sufficient to fully recover the spatial extent of bundles. All bundles reconstructed using the masking priors and the enhanced FOD show an increase in both bundle coverage and quality of the fanning. For the PT, the lateral portions of the pre- and post-central gyri are better covered while they are almost non-existent with standard PFT. Similarly, the CC has the fanning fully covering the pre-central and post-central gyri. As for the AF, the spatial extent in the frontal and temporal lobes is increased, but streamlines are also reaching more laterally into the gyri. The quantitative impact of the improved fanning directly impacted the volumes of the reconstructed bundles, as seen in the first column of Table [5.1](#).

### 5.3. RESULTS

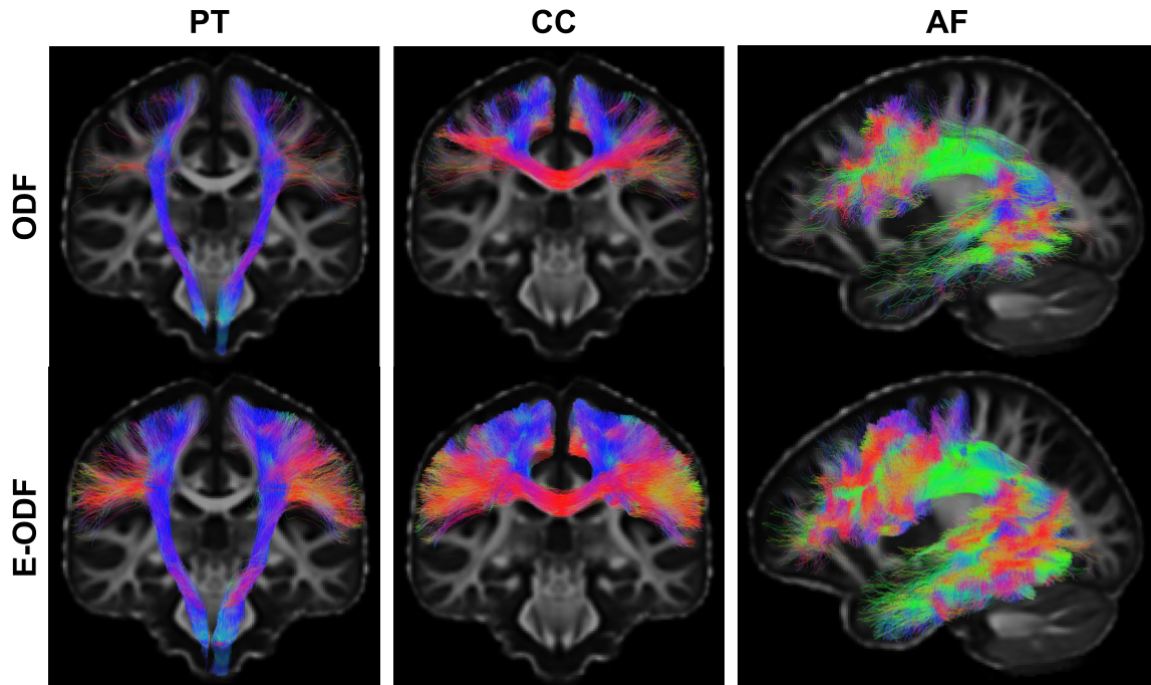


figure 5.7 – Examples of resulting PFT and BST algorithms for a single subject. The first row shows results from standard tracking of the PT, CC and AF, while the second row shows results from the proposed bundle-specific tracking. Of note, at no point tracking is allowed outside of the initial white matter tracking mask. It involves that no streamline travel in gray matter or CSF and the impression of “going over the sulci” which can be observed in the second row is due to coronal view lacking perspective.

### 5.3. RESULTS

#### Left Pyramidal Tract

	Local model	Mask	Bundle vol. ( $cm^3$ )	Dice index	Average length (cm)	% of valid streamlines	Efficiency (%)
1.	FOD	WM	$9.80 \pm 3.42$	$0.63 \pm 0.12$	$12.58 \pm 0.38$	$0.29 \pm 0.20$	$1.30 \pm 0.92$
2.	FOD	BOI	$14.70 \pm 4.00$	$0.71 \pm 0.10$	$12.66 \pm 0.32$	$0.67 \pm 0.38$	$4.01 \pm 2.45$
3.	E-FOD	BOI	<b><math>27.95 \pm 6.24</math></b>	<b><math>0.86 \pm 0.06</math></b>	$12.16 \pm 0.30$	<b><math>3.92 \pm 1.95</math></b>	<b><math>11.73 \pm 6.49</math></b>

#### Right Pyramidal Tract

	Local model	Mask	Bundle vol. ( $cm^3$ )	Dice index	Average length (cm)	% of valid streamlines	Efficiency (%)
1.	FOD	WM	$6.56 \pm 2.74$	$0.51 \pm 0.15$	$12.03 \pm 0.32$	$0.14 \pm 0.10$	$0.58 \pm 0.42$
2.	FOD	BOI	$11.08 \pm 3.42$	$0.64 \pm 0.13$	$12.07 \pm 0.26$	$0.38 \pm 0.22$	$2.21 \pm 1.32$
3.	E-FOD	BOI	<b><math>25.27 \pm 5.33</math></b>	<b><math>0.83 \pm 0.06</math></b>	$11.59 \pm 0.24$	<b><math>2.42 \pm 1.19</math></b>	<b><math>7.26 \pm 3.95</math></b>

#### Corpus Callosum

	Local model	Mask	Bundle vol. ( $cm^3$ )	Dice index	Average length (cm)	% of valid streamlines	Efficiency (%)
1.	FOD	WM	$25.78 \pm 6.58$	$0.65 \pm 0.11$	$11.25 \pm 0.37$	$0.72 \pm 0.33$	$2.89 \pm 1.39$
2.	FOD	BOI	$49.51 \pm 7.40$	$0.78 \pm 0.07$	$11.49 \pm 0.31$	$2.68 \pm 0.85$	$14.09 \pm 5.17$
3.	E-FOD	BOI	<b><math>87.23 \pm 7.71</math></b>	<b><math>0.86 \pm 0.06</math></b>	$11.53 \pm 0.26$	<b><math>12.20 \pm 3.46</math></b>	<b><math>30.32 \pm 9.96</math></b>

#### Left Arcuate Fasciculus

	Local model	Mask	Bundle vol. ( $cm^3$ )	Dice index	Average length (cm)	% of valid streamlines	Efficiency (%)
1.	FOD	WM	$54.21 \pm 9.49$	$0.70 \pm 0.08$	$10.61 \pm 1.30$	$1.23 \pm 0.45$	$4.92 \pm 1.94$
2.	FOD	BOI	$56.27 \pm 10.06$	$0.70 \pm 0.08$	$10.78 \pm 1.26$	$1.48 \pm 0.55$	$7.99 \pm 3.28$
3.	E-FOD	BOI	<b><math>110.03 \pm 10.23</math></b>	<b><math>0.75 \pm 0.06</math></b>	<b><math>12.29 \pm 0.72</math></b>	<b><math>9.74 \pm 2.80</math></b>	<b><math>32.26 \pm 11.14</math></b>

#### Right Arcuate Fasciculus

	Local model	Mask	Bundle vol. ( $cm^3$ )	Dice index	Average length (cm)	% of valid streamlines	Efficiency (%)
1.	FOD	WM	$35.89 \pm 9.97$	$0.58 \pm 0.13$	$9.35 \pm 2.44$	$0.77 \pm 0.36$	$2.71 \pm 1.60$
2.	FOD	BOI	$37.99 \pm 10.26$	$0.59 \pm 0.13$	$9.55 \pm 2.37$	$0.95 \pm 0.44$	$4.62 \pm 2.69$
3.	E-FOD	BOI	<b><math>81.84 \pm 12.99</math></b>	<b><math>0.80 \pm 0.06</math></b>	<b><math>11.60 \pm 0.82</math></b>	<b><math>5.95 \pm 2.46</math></b>	<b><math>18.89 \pm 9.15</math></b>

tableau 5.1 – Quantitative results are shown separately for each bundle of interest from three experiments, measures of the 39 validation subjects were averaged together. Values for average and standard deviation are formatted to simplify interpretation. All measures, except average length, are improved by our approach when compared to classical tractography (approach 1 and 2), the difference is statistically significant ( $p < 0.01$ ). Average length remains stable across method, however it was significantly higher for both arcuate fasciculus.

### 5.3. RESULTS

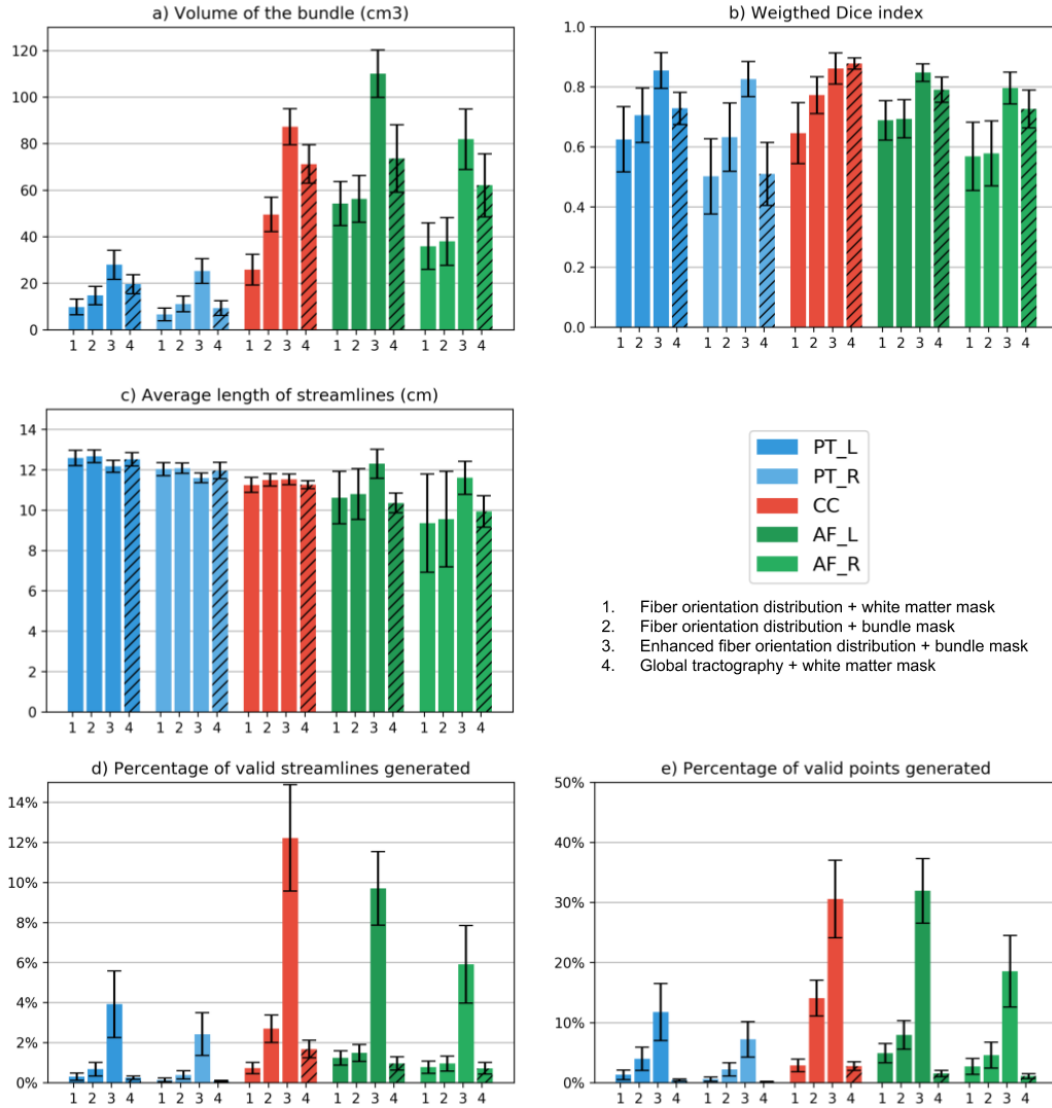


figure 5.8 – Quantitative measures for each BOI across the 3 tracking algorithms. Results from global tractography is added to provide more context, but comparai-son is partly “unfair” as many variables cannot be controlled in the same way. The Results are shown for probabilistic tracking with PFT. Graphs for deterministic and probabilistic versions of the algorithms with and without PFT are also available in the [BST, anatomical definition](#) Section

### 5.3. RESULTS

Figure 5.8 shows the trends for each measures of the analysis, where each graph represents the changes of all bundles across the three sets of tractography parameters. The impacts are not the same for each bundle, as the shape, size and the underlying anatomy are inherently different. Error bars in the histograms represent the average values and their standard deviation across 39 subjects. As the histograms can be hard to compare, values are also shown in Table 5.1.

In Figure 5.8a), an increase in volume for all bundles is measured when the priors are used. The AF is the only bundle with no significant increase between the tracking using no *a priori* masking and the tracking using only the *a priori* masking, however the volume significantly increases when enhanced FOD are used. The standard deviation, for methods 2 and 3, in proportion to the average, is decreasing. The volume of the reconstructed PT using method 3 is approximately 400% of the volume obtained from classical tracking, 300% for the CC and 200% for the AF.

The reproducibility across datasets was measured by the Weighted-Dice coefficient. Tractography without priors poorly performed with values between 0.5 (PT right) and 0.7 (AF left). In Figure 5.8b), an increase in overlap for all BOI is observed when E-FOD are used, with Weighted-Dice values all more or less 0.8. Similarly to the volume, the AF are the only bundles with no significant increase between methods 1 and 2, while the overlap significantly increases when enhanced FOD are used. For methods 2 and 3, the standard deviation of the Weighted-Dice coefficient is decreasing as the average coefficient is increasing. For all bundles, except the left AF, the Weighted-Dice increases by 20% when method 3 is used instead of the classical tracking.

The average length and standard deviation was computed to evaluate the shape coherence of each bundle. Globally, the average length of most BOI remained similar for each method, as seen in Figure 5.8c). However, for the AF, going from methods 2 to 3, the average length increased and the standard deviation decreased. This can be observed more easily in Table 5.1. As mentioned earlier (Figure 5.7), this is due to streamlines reaching further inside the lateral gyri.



### 5.3. RESULTS

Finally Figure 5.8d) and e) and the last two columns in Table 5.1 show measures of efficiency. The use of adapted tracking masks (method 2) is not remarkably influencing the percentage of valid streamlines. However, integrating enhanced FOD (method 3) significantly increases efficiency. A tractogram with one million streamlines, each streamline with more or less one hundred points, would need one hundred million iterations. However, not all streamlines have the same number of points depending on their length. When adapted tracking masks (method 2) are used, most invalid streamlines are quickly stopped and thus a large number of iterations avoided. Also, it is important to note that the percentage of valid streamlines and efficiency are improved but that they are still under 15% and 40% respectively, which leaves plenty of room for further improvements.

Similar trends can be observed in the [BST, anatomical definition](#) Section for probabilistic tractography (without particle filtering). Bundle-specific tractography also improves reconstruction when using deterministic algorithms, the difference between classical deterministic and deterministic bundle-specific tractography was statistically significant ( $p < 0.01$ ). This shows that most algorithms using FOD as an input need orientational priors to better reconstruct bundle of interest.

Results from MITK global tractography [[Neher et al., 2012](#)] were overall superior to all classic tractography algorithms used for the experiments, as seen in [BST, anatomical definition](#) Figures A.2,A.3,A.4,A.5. However, bundle-specific probabilistic tractography always generate bundles with higher volume and higher reproducibility across subjects compared to global tractography. The orientational prior used with probabilistic tractography (with or without PFT) achieve better results while decreasing computational needs.

## 5.4 Discussion

Our bundle-specific tractography algorithm is designed to address the poor spatial extent challenge and the sensitivity-specificity trade-off of tractography. Contributions are : i) novel bundle-specific anatomical and orientational priors directly incorporated in the tractography process to boost streamline tracking in hard-to-track regions, ii) results that show increased reproducibility, sensitivity, and specificity of PT, CC, AF bundle reconstructions in the complex centrum semiovale region, and iii) a new BST algorithm with increased efficiency and reduced overall computation time. The main methodological innovation is the novel local fiber orientation distribution enhancement methodology based on white matter bundle priors. These additions were shown to yield better spatial coverage of the pyramidal tract, corpus callosum, and arcuate fasciculus, increasing the quality of the extent of fanning reaching the cortex, helping to accurately represent the shape of each fascicle. Increase in volume is not a direct measure of the quality of a bundle reconstruction. The use of a strict anatomical definition based on the literature is fundamental to correctly address this concern. The definition was designed to reduce as much as possible the number of false positives, meaning that significant increase in volume cannot be accounted by an increase in outliers, spurious streamlines or contributions from other bundles. As mentioned earlier, tracking is never allowed outside of the initial white matter tracking mask. Hence, increase in volume cannot come from “*new voxels*” added by our algorithm. The increase in volume was observed in regions that initially lack fanning reaching the cortex, which is a common difficulty for classical tractography [Girard et al., 2014; St-Onge et al., 2018]. It is important to note that the FOD enhancement affects the mathematical properties of the FOD. The E-FOD should not be used to compute FOD metrics such as the average fiber density (AFD) [Raffelt et al., 2017] or even number of fiber orientations (NuFO) [Tournier et al., 2008] since lobes amplitude changed and no longer agree with the DWI signal.

## 5.4. DISCUSSION

### 5.4.1 Template creation needed for Bundle-Specific Tractography

The creation of an appropriate template of streamlines is crucial for the proposed method, as mentioned in the Methods Section. The input bundle of interest of the template needs to be strictly segmented and cleaned of spurious streamlines, while still representing all the potential anatomical variability across subjects. We acknowledge that anatomical definition of WM bundles can vary among experts and that there exist several open controversies in the WM neuroanatomy literature [Forkel et al., 2014; Hau et al., 2017; Meola et al., 2015].

However, having a consensus on the definition of the bundle is not necessary to showcase the potential of our Bundle-Specific Tractography. Neuroanatomists can debate and converge on the precise definitions of anatomical bundles, which can then be adjusted into our BST method.

Moreover, the nonlinear deformation of the template to the subject’s native space is done under the assumption that ANTS performs an adequate registration in-between FA maps. In the case of a different study, with pathology or large brain alterations, such as a tumor, cyst, severe atrophy, or missing brain regions, the quality of the alignment would need to be guaranteed manually. Using a tailored template based on a specific cohort enables our proposed BST method to be fully automatic without relying on manually placed ROIs to enforce certain directions and ending regions. Pathological brains already raise numerous challenges for tissue segmentation, registration, classical tractography and analysis, which all require case-by-case decisions and tuning. For example, WM lesions for Alzheimer’s disease, multiple sclerosis, traumatic brain injuries do not have the same underlying causes and tractography could sometimes be allowed to go through them or not. When all conditions are met and potential pitfalls related to the pathology are accounted for, BST can yield interesting results and can have positive impacts for challenging datasets, as illustrated on the fornix reconstruction in an aging study [Rheault et al., 2018a].

As atlases of streamlines become more widely available, such as the one of Wasserthal et al. [2018a]; Yeh et Tseng [2011], different research groups will be able to

## 5.4. DISCUSSION

use these without having to build their own. As an attempt to facilitate its use and guarantee the quality of the anatomical priors for well-known pathways, a WM atlas dedicated to the BST has been designed. This is a step in the right direction to include anatomical priors to improve tractography and address some of the challenges raised in [Maier-Hein et al. \[2017\]](#). To facilitate future investigations and help other researchers to use our method, a *Nextflow* pipeline [[Di Tommaso et al., 2017](#)] has been developed. By providing configuration files, a lightweight WM atlases of bundles and a singularity environment [[Kurtzer et al., 2017](#)], this pipeline will simplify the usage of our method. Also, researchers with specific needs can easily customize the pipeline and the configuration file and obtain exactly what they need.

### 5.4.2 Bundle volume and spatial coverage

The priors on the regions of interests and local orientations noticeably improve the virtual dissection results by increasing the total volume and the fanning extent. FOD enhancement helps to overcome tracking difficulties in the crossing regions, allowing more streamlines to respect the anatomical definition of the bundle of interest (see [BST, anatomical definition](#) Section). Difficulties encountered in the crossing region would have required a higher number of seeds per voxel to obtain an appropriate reconstruction for the classical method, however this approach does not ensure to fully capture the spatial extent. When FODs are enhanced to support the main directionality of a bundle of interest, crossing regions become less disorienting during the tractography, thus increasing the number of streamlines to be considered during segmentation. All this while being more efficient, in term of valid streamlines after segmentation.

As the fanning component of the bundles represent the majority of their volume, improved fanning has a quantitative impact on the volume. For example, the fanning of the PT and CC occurs in the 3-way crossing region (Figure [5.1](#)), traditionally resulting in a small preponderance of streamlines reaching lateral portions of the primary motor and somatosensory cortex. The priors largely impact the volume of the bundles (first column of Table [5.1](#)). The use of *a priori* seeding tracking masks increases volume as particle filtering tractography enforces streamline termination in

## 5.4. DISCUSSION

GM regions. The usage of enhanced FOD increases the volume even more by reducing the potential confusion in crossings, as picking a bad direction at each step quickly accumulates in terms of error.

### 5.4.3 Reproducibility

Bundles of interest considered in this project are known to have anatomical variability at the micro-scale, but at the macro-scale (on the whole bundle level) the AF, CC and PT are expected to cover the same regions (once in a common space) after the segmentation described in the [BST, anatomical definition](#) Section. We acknowledge that an identical coverage is not expected, meaning the weighted-dice should not be expected to be 1.0. However, poor spatial agreement (inter-subject) with the classical tractography method mainly comes from poor fanning, a typical solution to this problem is to seed more until the expected shape is obtained and fanning is adequate. The low weighted-dice value seen from classical method is influenced by the high variability of the fanning (CC, PT) or from the early-stop of streamlines in the frontal and temporal lobe (AF). The goal was not to obtain a value of 1.0. The goal was to reduce the variability in these regions as they were caused by tractography limitations and not explained by the underlying anatomy. Intra-subject reproducibility could have hidden this pitfall as a lack of fanning could be reproducible (intra-subject) and would not have reflected if the expected anatomical shape was achieved.

### 5.4.4 Computational Performance

Immediately stopping any streamline getting out of the bundle of interest region using a priori imposed dilated masks directly reduces the number of tractography iterations and saves computation time without altering the resulting dissection. Such stopping criteria also decreases the rate of false positives, as streamlines must follow a reasonable path to be considered valid. The efficiency of BST is also increased since streamlines follow the main direction of the BOI instead of leaving prematurely the tracking mask without reaching ROI endpoints. By following the appropriate direction in the crossing, the streamlines have more chance to respect all the bundle definitions, resulting in a higher amount of valid streamlines.

#### 5.4. DISCUSSION

Performance is not just about number of streamlines. The increase in performance combined with the increase in volume and overlap between subjects (Table [5.1](#)) shows that using the proposed method helps to reach stable results more rapidly, while improving the quality of the bundle of interest.

## 5.5. CONCLUSION

### 5.4.5 Future Work

Since the output of the proposed method is in the form of FOD in a widespread format (spherical harmonics) and binary masks, any algorithm that accepts such input could be easily adapted. Other forms of tracking (Dipy [Garyfallidis et al., 2014], FSL [Behrens et al., 2007], MRtrix [Tournier et al., 2012], Real-time tractography [Chamberland et al., 2014], even Global tractography [Christiaens et al., 2015; Krehner et al., 2008; Reisert et al., 2011]) could be tried. Algorithms designed to add more *a priori* could also be tried alongside enhanced-FOD such as surface-enhanced tractography [St-Onge et al., 2018] and microstructure-informed tractography (AxTract) [Girard et al., 2015]).

## 5.5 Conclusion

We developed a new bundle-specific tracking (BST) algorithm incorporating novel anatomical orientational priors directly into the streamline tracing process to address the poor-spatial-extent challenge and sensitivity-specificity trade-off of WM bundle reconstruction using tractography. This fully automatic method exploits information extracted from a template of streamlines of bundles of interest to enhance the local modeling in the desired direction using the *a priori* fiber orientation distribution. It was shown that the proposed approach improves the spatial coverage and increases the quality of the fanning in crossing regions, while reducing computational need. Since a stable volume and sufficient anatomically valid streamlines are reconstructed faster than with classical techniques, this method helps the exploration of structural connectivity with more confidence and less biases in white matter bundles metrics. This could have a positive impact on the neurosciences community using dMRI tractography tract-based and connectivity-based analyses.

“That which can be asserted without evidence, can be dismissed without evidence.”

---

— Christopher Hitchens, *God Is Not Great*.



# Discussion

## Perspectives en lien avec cette thèse

### Création d’atlas de faisceaux

Les algorithmes *RecobundlesX* et *BST* présentés aux Chapitre 4 et 5 reposent sur un système de patron représentant la position et morphologie générale d’un ou plusieurs faisceaux. Étant donné la variabilité des définitions anatomiques, comme mentionnés à la Section 2.3.2 et au Chapitre 3, ces patrons peuvent varier selon l’utilisateur et même selon l’utilisation désirée. Il est impératif pour notre domaine de recherche de formaliser, en tant que communauté, certaines définitions anatomiques et de s’entendre sur les protocoles de segmentation de certains faisceaux. Sans ce type d’effort, les méthodes basées sur des *a priori* seront vouées à être adaptées pour chaque projet sur une base individuelle et resteront donc complexes, ou carrément impossibles à utiliser dans un contexte clinique. Les définitions anatomiques qui devront être mises en place ne seront pas nécessairement optimales pour tous, mais seront une référence cruciale pour les applications cliniques. L’obtention d’un consensus entre les experts en neuroanatomie et en tractographie sera probablement très difficile.

À ce jour il n’y a toujours pas de consensus sur le pré-traitement de l’IRMd, encore moins pour la tractographie. Le *2015 ISMRM WM challenge* [Maier-Hein et al., 2017] a recueilli des soumissions de plus de 70 équipes internationales, celles-ci présentaient une variété d’étapes de traitement de données et une vaste distribution de qualité de résultats. Par contre, les discussions liées aux définitions anatomiques ne sont pas indépendantes du pré-traitement et de la tractographie et donc il serait nécessaire d’intégrer le tout. Les efforts liés à la conception du livre *Handbook of Tractography*

## DISCUSSION

vont probablement permettre d'attaquer bon nombre de ces questions et unifier en quelque sorte les recommandations faites pour les applications cliniques.

## Validations de la microstructure et la tractographie

Jusqu'à maintenant la limitation majeure de la tractographie est son manque d'adaptation selon sa position et le manque de «*connaissances*» de l'anatomie sous-jacente. En effet, la tractographie n'est, dans la vaste majorité des cas, qu'une séquence d'itérations aveugle dans un champ de vecteurs, aucune information de la provenance, de la destination ou même des routes connues n'est injectée dans les reconstructions. Les avancées rapides en intelligence artificielle vont permettre le développement d'algorithmes incorporant une multitude d'*a priori* et ainsi faciliter la reconstruction de tractogrammes valides. Par contre, il est essentiel que ces *a priori* soient basés sur des données d'une qualité exceptionnelle, il est donc important de valider ceux-ci grâce à des méthodes de validation, telles que celles mentionnées à la Section 1.4. La fusion des connaissances provenant de la validation et des meilleurs algorithmes classiques devra être évaluée de façon rigoureuse pour éviter que les algorithmes d'intelligence artificielle reproduisent les biais présents en ce moment. La complexité du cerveau est trop grande pour que nous puissions créer manuellement un ensemble de règles variant selon le contexte et la position.

## Évaluation des sources de variabilité

Au cours des dernières années, de nombreux projets se sont attaqués à la quantification de la variabilité liée à l'IRMd et la tractographie. La variabilité dans le temps (test-retest), entre scanners d'un même fabricant, entre fabricants de scanners, les schémas d'acquisitions et les paramètres de pré-traitement ou de post-traitement sont au cœur de cet effort de quantification. Mais certains poussent un peu plus loin ; depuis longtemps la communauté a ignoré l'importance de la distribution de la population à l'étude. Les ratios entre sexe, âge, dominance manuelle, minorités ethniques ou l'incorporation de population ayant des anomalies génétiques, dérèglements hormonaux et même pathologies doivent être plus longuement réfléchis. Par exemple, pendant plusieurs centaines d'années l'étude du cerveau par dissection était focalisée presque

## DISCUSSION

exclusivement sur des populations soit saines, mâles et caucasiennes. La recherche qui découle de l'étude de variabilité, autant algorithmique que populationnelle, montre que celle-ci est sous-estimée et peut avoir des conséquences imprévues sur la réplicabilité des résultats. Il est donc important de rester, pour le moment, sceptique des affirmations et interprétations excentriques issues de la tractographie. L'utilisation de vaste volume de données issues de bases de données telles que le Human Connectome Project [Van Essen et al., 2013] ou la UKBioBank [Sudlow et al., 2015] pourront permettre de pallier à certaines de ces limitations.

L'objectif à long terme est de démocratiser l'IRMd et la tractographie en clinique. La création d'empreintes de connectivité robustes permettrait d'aider au diagnostic ou de comprendre divers processus cognitifs. Les outils ouverts de *Scilpy* et *Dipy*, l'outil de visualisation libre *MI-Brain*, les pipelines *Nextflow* et les environnements *Singularity* auxquels j'ai participé vont permettre aux méthodes développées au cours de cette thèse de survivre à l'usure du temps.

## Perspectives en lien avec la communauté

Les conférences annuelles permettent d'observer les tendances prises par la communauté et évaluer à quoi va ressembler le paysage scientifique des années à venir. Cette section n'est qu'une tentative de spéculation sur ces tendances, mais aussi un commentaire général sur ce qui est, à mon avis, nécessaire pour notre communauté.

## Méthodes et algorithmes

### Visualisation

L'une des principales caractéristiques du domaine de la tractographie est sa forme unique, composée de lignes en 3 dimensions et colorées, qui permet instantanément d'identifier celle-ci. À chaque nouvelle publication, une nouvelle façon de visualiser les données et résultats est utilisée, les unes toujours plus complexes que les autres. Il est de mon humble avis que beaucoup de visualisation de la connectivité de la matière blanche se fera à l'aide du cortex. La visualisation de la parcellisation de la matière grise se fait grâce à des rendus de la surface du cortex ; pourquoi pas la parcellisation

## DISCUSSION

de la matière blanche [St-Onge et al., 2018]? Des cartes de chaleurs de différences morphologiques, de la variabilité inter-acquisitions, de l'impact de pathologies sur la matière blanche, les intersections et goulots d'étranglement, le tout projeté au cortex.

*Résumé : Une meilleure intégration de la visualisation de la matière blanche et de la surface corticale sera courante.*

## Parcellisation

Le terme *parcellisation* est généralement associé à la matière grise, mais la classification de la matière blanche profonde peut aussi porter le même nom. La prolifération des parcellisations corticales en neuroimagerie est impressionnante, et chacune est meilleure que la suivante (ou aucune ne l'est réellement?). Un outil «*grand public*» tel que FreeSurfer a démocratisé la parcellisation *Desikan-Killiany*, ce qui fait que celle-ci est maintenant extrêmement répandue. La classification des faisceaux de la matière blanche va potentiellement suivre un sort similaire. Une situation d'oligopole va se développer et quelques outils ouverts très populaires vont se partager le marché. Malgré l'utilisation de définitions de faisceaux, anatomiquement controversées et truffées de limitations, certains outils vont dominer la tâche de parcellisation.

*Résumé : L'accessibilité et la facilité d'utilisation seront au coeur de la compétition pour la parcellisation de la matière blanche la plus populaire.*

## Biophysique

Bien que la modélisation locale ait été largement exclue des travaux présentés dans cette thèse, elle est bel et bien la base de la tractographie (pour l'instant). La modélisation par compartiments est une méthode d'avenir pour ceux qui préfèrent la microstructure plutôt que la tractographie. Par contre, la tractographie a aussi grandement besoin d'informations locales sur la microstructure pour compléter l'information d'orientation. Simplement itérer dans un champ de vecteurs ne suffira jamais pour surpasser les défis de la tractographie liés aux faux positifs. L'incorporation à la tractographie d'information de microstructures en encourageant la cohérence spatiale (indice de dispersion, diamètre axonal, myélinisation, etc.) sera cruciale. Une telle approche est extrêmement rigide et pourra difficilement remplir le volume entier

## DISCUSSION

de la matière blanche. Par contre, il pourrait s'agir d'une approche visant plutôt à caractériser la connectivité à une échelle plus macroscopique, faciliter la classification de certaines «*autoroutes*» ou visualiser la variation de certains biomarqueurs dans la matière blanche.

*Résumé : Le temps des modèles locaux et de la tractographie basée uniquement sur l'orientation locale et contrainte par de simples règles géométriques est révolu.*

## Neuroanatomie

Beaucoup d'algorithmes de neuroimagerie sont développés uniquement en temps que solution pour un problème spécifique à un jeu de données, à un laboratoire ou même pour un projet. Ce manque de consultation entre les groupes de recherche encourage la prolifération (et la mort rapide) de plusieurs algorithmes. Au-delà des avancées en termes d'algorithmique, une réflexion s'amorcera sur la réalité anatomique sous-jacente qui est liée à l'algorithme. Par exemple, la segmentation de tissus sains est plus qu'une classification entre blanc, gris et noir. Ce que représentent ces tissus est crucial pour la tractographie. Est-ce qu'une division en 3 classes représente réellement toute la complexité du cerveau ? Est-ce qu'une *streamline* peut traverser le volume partiel d'un noyau si celle-ci ne fait que l'effleurer parallèlement à sa surface ? Est-ce que la tractographie doit ou non traverser un tissu pathologique ? La classification par tissus du cerveau (sans même mentionner la tractographie) est encore un problème difficile et sans réponse claire, et pourtant il s'agit d'une entrée pour la tractographie.

*Résumé : Chaque étape précédant la tractographie va être revisitée pour intégrer plus d'a priori anatomiques*

## Mathématiques

Les mathématiques sont dissimulées un peu partout dans le traitement de données lié à la tractographie. Il est évident que l'optimisation des séquences d'acquisition et la modélisation locale auront toujours besoin de mathématiques. La portion encore «*inexplorée*» du spectre du traitement de données est dans le parcours du champ de directions. La navigation de champ de directions n'est pas un problème unique à la tractographie et de nombreuses inspirations venant d'autres communautés seront les

## DISCUSSION

bienvenues. Qu'il s'agisse de mécanique des fluides, de modélisation de flot d'air en météorologie, la modélisation d'essaim ou de contraintes de mouvements dans l'espace. Le tout dans un scénario chaotique où le moindre changement aux configurations initiales a une influence majeure.

*Résumé : Les algorithmes de tractographie de plus en plus complexes vont devenir de plus en plus courant.*

## Statistiques

Encore à ce jour les analyses de résultats liés à la tractographie sont très peu répliquables et difficilement généralisables. La complexité de l'acquisition, du traitement, de la modélisation et de la reconstruction par tractographie ne font qu'ajouter des variables difficiles à prendre en compte lors de l'analyse de résultats. Les efforts récents pour *harmoniser* les données et noter les sources de variabilité sont un pas dans la bonne direction. Beaucoup d'efforts supplémentaires seront nécessaires pour rapporter et corriger ces sources de variabilité. Qu'il s'agisse de la variabilité *test-retest*, inter-scanner, inter-fabrique, inter-algorithme de pré-traitement, variabilité à l'évaluation des tests cognitifs, ces sources peuvent et devront être prises en compte lors des analyses statistiques.

*Résumé : Les analyses vont commencer à prendre en compte la variabilité expérimentale dans leur présentation des résultats.*

## Approches en traitement de données

### Puissance de calcul

L'augmentation de la capacité de calcul et de l'espace disque facilite grandement le développement rapide de méthodes de recherche. Il fut un temps où un volume de diffusion à 12 directions en 2mm x 2mm x 2mm ne pouvait entrer en RAM, un peu comme une texture de jeu vidéo ne pouvait dépasser 32Ko. De nos jours, une simple étape de débruitage nécessite 16Go de RAM pour un volume de diffusions à 270 directions en 1.25mm x 1.25mm x 1.25mm. Alors qu'un jeu vidéo utilise des textures en 4K et des milliers de polygones sont visibles dans une scène typique. Le temps de rendu d'une seule image présente dans un film était de 8h en 1990, 8h en

## DISCUSSION

2000, 8h en 2010 et probablement encore 8h en 2020. Bien certainement, la qualité du résultat augmente aussi, mais l'approche algorithmique et la durée typique des étapes individuelles ne varient que très peu. Les nouveaux algorithmes continueront de se complexifier à mesure que la puissance de calcul augmentera.

*Résumé : Les gigantesques bonds en avant, en temps d'exécution, seront atteints par des avancées mathématiques plutôt que matérielles.*

### Calcul distribué

L'accumulation de grandes bases de données a été facilitée au cours des dernières années. Par contre, le traitement de ces bases de données est beaucoup plus difficile. La limitation en temps de traitement est contournée par l'utilisation de superordinateurs et de plateformes infonuagiques à grande échelle. La limitation n'est plus autant le temps CPU ou la RAM, mais bien le stockage des données traitées qui peuvent être 10-100 fois plus volumineuses que les données originales. Les considérations environnementales du stockage à long terme de données dormantes vont peut-être encourager la communauté à partager les données autant que possible, qui sait ? La démocratisation de *pipelines* de traitement 100% reproductible [Theaud et al., 2019] et l'exécution minutieuse du traitement et de l'analyse va permettre un essor de la quantité de données traitées. Pour le stockage à long terme, il sera plus simple de ne conserver que le *pipeline* et les données initiales. Tout comme une prolifération des méthodes algorithmiques s'est observée pour toutes les étapes de traitement pour ensuite se réduire à quelques méthodes plus populaires et intuitives d'utilisation, le même phénomène s'observera à l'échelle des *pipelines*.

*Résumé : L'accessibilité au calcul distribué va se faciliter et de nombreux pipelines, de qualité variable, vont faire surface.*

## L'apprentissage machine, inévitable ?

### Segmentation d'images et de faisceaux

Le domaine de la neuroimagerie est loin derrière l'état de l'art en apprentissage machine (ML). Un énorme effort de *traduction* sera mis en place pour appliquer des méthodes bien connues dans d'autres domaines et les appliquer à nos besoins. Les

## DISCUSSION

travaux en cours en ce qui concerne les images volumiques (classification de tissus, segmentation de lésions, approche par focus pour les pathologies, catégorisation de participants) sont déjà bien avancés. Une certaine tendance à tout redéfinir et présenter des approches déjà connues comme innovatrices, simplement parce qu’elles sont présentées en lien avec la neuroimagerie, est observable. Ces approches vont continuer de se répandre rapidement due à la similarité entre une image volumique (T1, IRM<sub>d</sub> IRM<sub>f</sub>) et les images 2D ou les vidéos. Par contre, la conversion de ces approches à l’objet 3D que sont les *streamlines* n’est pas triviale. La classification de faisceaux ou le filtrage de tractogramme sont des défis particulièrement intéressants. De nombreuses approches classiques existent dans l’espace des *streamlines*, mais encore peu proviennent du domaine du ML. Pour ce qui est des images, le ML est déjà incontournable pour plusieurs tâches (bien qu’encore marginal dans le monde clinique). La représentation par *streamlines* va suivre, quoique beaucoup plus lentement.

*Résumé : C’est dans les approches spécifiques aux streamlines que résident l’innovation en ML pour notre communauté.*

## Reconstruction par tractographie

Il s’agit probablement de secteur le plus prometteur de l’AI pour notre communauté. Les algorithmes de tractographie intègrent l’information issue d’une séquence incroyablement complexe d’étapes. Diminuer la dépendance sur ces étapes de pré-traitement, harmoniser les données (inter-scanner, inter-fabrique, etc.), éviter la création de règles morphologiques et anatomiques explicites tout en procurant une accélération impressionnante, tout ceci est possible avec la tractographie utilisant la puissance du ML. Il est improbable qu’un algorithme faisant tout ça apparaisse du jour au lendemain, mais tout comme dans le domaine des images il est possible de juxtaposer des modèles et ainsi réduire le nombre total d’étapes un morceau à la fois. Par exemple, les étapes d’harmonisation, de débruitage et de ré-échantillonnage peuvent être combinées. Tandis que pour la modélisation locale, l’injection d’*a priori* de forme, le traçage de *streamlines* et le filtrage de résultats anatomiquement aberrants (non plausibles) peuvent être combinés et utilisés juste après la première étape.

*Résumé : La simplification des étapes de traitement pour obtenir des résultats équivalents, au départ, est la voie d’entrée de l’apprentissage machine dans notre domaine.*



## DISCUSSION

### Remise en question de la «vérité»

Une approche typique pour l'apprentissage machine est d'*apprendre* à partir de plusieurs exemples issus d'une variété d'algorithmes classiques et de variations d'hyperparamètres. Par contre, certains de ces algorithmes classiques sont répandus pour leur facilité d'utilisation et non pour la qualité de leurs résultats. Il n'est pas difficile d'imaginer qu'un algorithme de ML surpasse de tels résultats, mais si la mesure d'évaluation est basée sur ce même résultat : comment s'en rendre compte ? Un exemple concernant les faisceaux provient des dissections virtuelles manuelles extrêmement minutieuses qui seront toujours limitées par l'approche utilisant des régions d'intérêt voxélique. Donc beaucoup de *streamlines* aberrantes sont toujours présentes dans ces données, cachées au travers d'un faisceau en apparence *propre*. Est-ce vraiment la cible optimale à reproduire pour une nouvelle méthode de ML ? Mais comment quantifier quelque chose qui surpasse sa propre cible ? De façon similaire au domaine de l'analyse photographique, qui a dû développer de nouvelles mesures permettant de mieux représenter la vision humaine ou les impressions qualitatives d'un observateur, le domaine de la neuroimagerie va devoir faire de même. Qu'il s'agisse de la représentation par voxels, surfaces ou *streamlines*, les mesures d'accord et de similarité vont devoir être redéfinies pour réellement représenter les connaissances abstraites de l'expertise humaine.

*Résumé : À mesure que les experts en apprentissage machine vont découvrir les limitations de la neuroimagerie, le concept même de «vérité terrain» va devoir être remis en question.*

### Démocratisation ou révolution

#### En clinique ?

L'utilisation en clinique se résume encore bien souvent à une recherche absolue du meilleur biomarqueur sans réellement se soucier des limitations cachées par la complexité du traitement de données. Une acquisition de 12 directions d'une résolution de 2mm x 2mm x 3mm, le modèle du tenseur de diffusion, suivi d'une tractographie déterministe, contrainte par un seuil d'anisotropie fractionnelle est considéré une utilisation de pointe en clinique et la simple utilisation du coefficient moyen de diffusion

## DISCUSSION

(ADC) est encore la seule utilisation quotidienne en clinique. Avec la démocratisation des outils de traitement (*pipelines*, outils et librairies) on pourrait penser que cette problématique disparaîtra. Sans la présence d'efforts actifs de collaboration entre le monde méthodologique et clinique, l'écart ne sera pas facilement réduit. L'attrait de la recherche fondamentale doit être balancé par des désirs concrets d'application. La communauté méthodologique doit s'efforcer de produire des résumés de leurs positions faisant le consensus et partager ceux-ci avec le monde clinique. Les séquences d'acquisition, modélisation locale, étapes de traitement, outils à utiliser, paramètres de tractographie, définition de faisceaux, méthodes d'analyses doivent être explicitement discutés ; la recherche fondamentale doit fermer cet écart. Sans ces efforts, le monde clinique va poursuivre ses efforts de recherche avec les méthodes intuitives et les outils déjà communs et connus. En ce qui concerne les efforts du côté clinique : Les limitations doivent être plus souvent reconnues, les études de réplication encouragées, les mauvaises pratiques méthodologiques dénoncées et les résultats négatifs publiés.

*Résumé : L'écart méthodologique entre recherche fondamentale et appliquée ne sera réduit que par une collaboration extrêmement étroite entre les deux mondes et non par la simple diffusion des connaissances.*

### Autres domaines ?

Cette thèse n'avait que pour seul organe d'intérêt le cerveau humain. Par contre, l'application des méthodes de tractographie, et plus généralement l'imagerie par résonance magnétique de diffusion s'étend à d'autres organes et espèces animales. Il est surprenant d'apprendre que les outils disponibles pour le traitement de données d'imagerie sont, dans la majorité des cas, plus développés pour le cerveau humain que pour d'autres organes ou animaux. Il est probable que cette recherche attire plus de chercheurs et donc génère plus de résultats. De plus, le cerveau est particulièrement facile à imager, immobile, relativement similaire d'un individu à l'autre et facile à isoler. Le coeur, quant à lui, est entouré par d'autres organes, sa morphologie et position est plus variable et bat rapidement en continu. Comparée à l'imagerie cérébrale, l'imagerie cardiaque est un véritable défi en termes d'acquisition. Par contre, une fois l'acquisition effectuée, la modélisation locale est plus simple (muscles fortement orientés/organisés sans croisement, en opposition à toute la complexité souvent

## DISCUSSION

méconnue de la matière blanche). Les défis qu'ont en commun ces domaines profitent mutuellement à tous, bien que les solutions soient parfois lentes à faire leur chemin d'un domaine à l'autre. Le partage des connaissances se faisant difficilement par le biais de publications méthodologiques.

*Résumé : Le domaine de l'IRMd aurait tout à gagner à faciliter le partage de connaissances liées aux différents organes et animaux et mettre en commun leurs outils.*

### Grand public ?

Tous ont une vague idée de ce qu'est la matière grise, ces plis à la surface du cerveau semblent avoir marqué l'imaginaire populaire. La matière blanche quant à elle semble moins populaire, probablement dû au manque de familiarité avec celle-ci. Bien peu de gens ont réalisé que la couverture de l'album *The 2nd Law* par *Muse* était en fait une représentation de la connectivité cérébrale (issue du Human Connectome Project). L'utilisation de plus en plus courante de la tractographie en clinique va peut-être mener à une connaissance accrue de ces méthodes. Par contre, il est plus probable que des avancées en visualisation scientifiques vont faire leur chemin jusqu'à la couverture de magazine scientifique ou même dans les médias grand public. Peut-être qu'un jour la matière blanche sera autant populaire et reconnaissable que la matière grise. En attendant, la meilleure approche est probablement de continuer les efforts de vulgarisation dans les médias et populariser l'imagerie cérébrale et les neurosciences en général.

*Résumé : La popularité de la tractographie auprès du public passe avant tout par la compréhension des techniques d'imagerie de base.*

“If you think this has a happy ending,  
you haven’t been paying attention.”

---

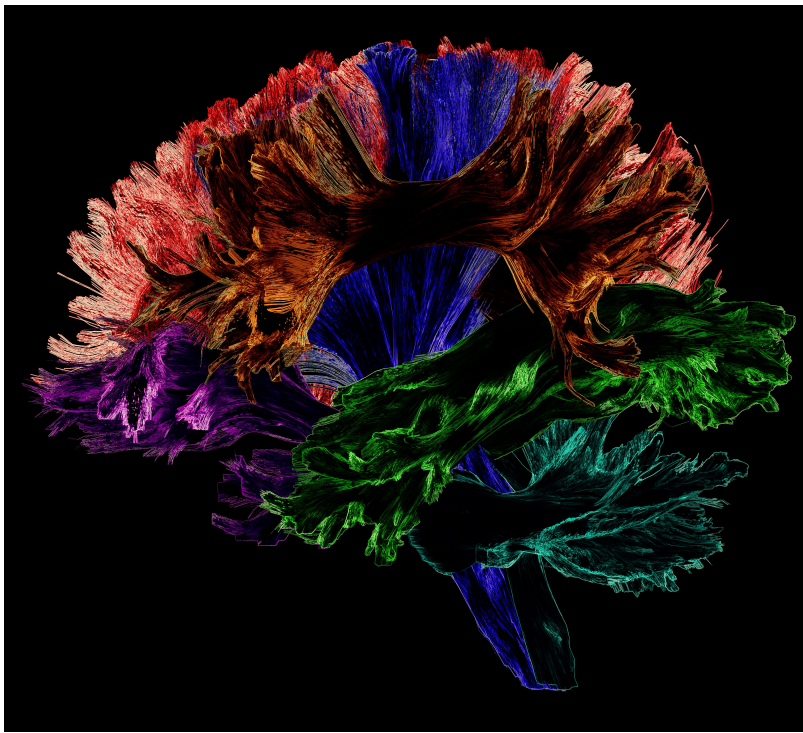
— Ramsay Bolton, Game of Thrones.

# Conclusion

L'imagerie cérébrale est un outil crucial pour les neurosciences modernes. Nous sommes présentement à l'âge d'or de l'imagerie par résonance magnétique de diffusion. La tractographie, l'imagerie microstructurale et autres type d'acquisition visant à modéliser ou comprendre la connectivité structurelle du cerveau gagne en popularité. Cette thèse a présenté une vision orientée vers la reconstruction et l'analyse de faisceaux de la matière blanche. Dans un premier temps, ce document a présenté des limitations et défis majeurs auxquels notre communauté fait face. Ensuite, cette thèse a introduit un cadre permettant une évaluation rigoureuse et démontré les problématiques liées au manque de reproductibilité dans les tâches de segmentations manuelles. Par la suite, une extension multi-atlas et multi-paramètres à l'algorithme *Recobundles* a été proposée. Cette nouvelle méthode permet d'améliorer et stabiliser la qualité des résultats de segmentation de tractographie. Finalement, une méthode de tractographie spécifique aux faisceaux a été présentée, ce nouvel algorithme permet de surmonter certaines limitations de la tractographie en injectant des *a priori* d'orientation. Le thème de mes travaux de doctorat était principalement axé sur la reconstruction et l'analyse de faisceaux. Elle a quelque peu délaissé les avancées liées à l'acquisition, la modélisation locale, l'intelligence artificielle ou les développements en lien avec la neuroanatomie. L'intégration de toutes ces avancées est loin d'être triviale et nécessitera un grand effort collaboratif multidisciplinaire.

## CONCLUSION

Cette thèse est la réalisation de 5 ans de travail, l'image ci-dessous a confirmé très tôt mon amour pour la recherche (et les faisceaux de la matière blanche). Elle m'a permis de rencontrer mon idole de jeunesse, Charles Tisseyre, de me donner un peu de matériel de vulgarisation sur ce que je faisais pour mes amis et ma famille. En plus d'avoir un beau souvenir de ma recherche sous forme d'un grand morceau d'aluminium brossé magnifiquement coloré qui est resté, et restera encore longtemps, dans le salon de mes parents.



**Description originale :** Nous avons tous une vague idée de la matière grise : ces replis à la surface du cerveau. Mais la matière blanche qui se cache en dessous est moins familière. Elle se compose d'une multitude de réseaux de câblages faits d'axones, qui sont les prolongements des neurones. Ces câbles sont recouverts de myéline, une substance blanche facilitant la circulation de l'information électrochimique. D'un cerveau à l'autre, ces réseaux d'axones présentent tous la même organisation, bien qu'ils aient chacun leurs caractères distinctifs. Grâce à l'imagerie médicale, et de puissants logiciels, on peut procéder à la dissection virtuelle d'un cerveau pour ne laisser apparaître que ses principales autoroutes nerveuses. *Image gagnante du prix du public pour le Concours Preuve par l'image de l'ACFAS 2016.*

“Anything one man can imagine,  
other men can make real.”

---

— Jules Verne, *Around the World in Eighty Days*.

# Publications résultant de cette thèse

## Articles de journaux

Premier auteur :

- **Rheault, Francois**, Etienne St-Onge, Jasmeen Sidhu, Klaus Maier-Hein, Nathalie Tzourio-Mazoyer, Laurent Petit, and Maxime Descoteaux. "Bundle-specific tractography with incorporated anatomical and orientational priors." *NeuroImage* : 186 (2019) : 382-398
- **Rheault, Francois** et al. "Tractostorm : The what, why and how of tractography dissection reproducibility." *Human Brain Mapping* (2020)
- **Rheault, Francois**, Guillaume Theaud, Philippe Poulin, Etienne St-Onge and Maxime Descoteaux. "Common misconceptions, hidden biases and modern challenges of dMRI tractography" *Neural Engineering* (2020)
- **Rheault, Francois**, Jean-Christophe Houde, and Maxime Descoteaux. "Visualization, interaction and tractometry : dealing with millions of streamlines from diffusion MRI tractography." *Frontiers in neuroinformatics* 11 (2017) : 42
- **Rheault, Francois**, Maggie Roy, Stephen Cunnane, and Maxime Descoteaux. "Bundle-specific fornix reconstruction for dual-tracer PET-tractometry." *BioRxiv* : 423459 (2018)

Co-auteur :

- Garyfallidis, Eleftherios, Marc-Alexandre Côté, **Rheault, Francois**, Jasmeen Sidhu, Janice Hau, Laurent Petit, David Fortin, Stephen Cunnane, and Maxime Descoteaux. "Recognition of white matter bundles using local and global streamline-based registration and clustering." *NeuroImage* 170 (2018) : 283-295
- Corrivetti, Francesco, Michel Thiebaut de Schotten, Isabelle Poisson, Sébastien Froelich, Maxime Descoteaux, **Francois Rheault**, and Emmanuel Mandonnet. "Dissociating motor-speech from lexico-semantic systems in the left frontal lobe : insight from a series of 17 awake intraoperative mappings in glioma patients." *Brain Struc-*



## CONCLUSION

- ture and Function 224, no. 3 (2019) : 1151-1165.
- Chenot, Quentin, Nathalie Tzourio-Mazoyer, **Francois Rheault**, Maxime Descoteaux, Fabrice Crivello, Laure Zago, Emmanuel Mellet et al. "A population-based atlas of the human pyramidal tract in 410 healthy participants." Brain Structure and Function 224, no. 2 (2019) : 599-612.
  - Mandonnet, Emmanuel, Guillaume Herbet, Sylvie Moritz-Gasser, Isabelle Poisson, **Francois Rheault**, and Hugues Duffau. "Electrically induced verbal perseveration : A striatal deafferentation model." Neurology 92, no. 6 (2019) : e613-e621.
  - Loit, Marie-Pierre, **Francois Rheault**, Etienne Gayat, Isabelle Poisson, Sébastien Froelich, Nanxi Zhi, Stéphane Velut, and Emmanuel Mandonnet. "Hotspots of small strokes in glioma surgery : an overlooked risk?." Acta neurochirurgica 161, no. 1 (2019) : 91-98.
  - Edde Manon, Guillaume Theaud, **Francois Rheault**, Bixente Dilharreguy, Catherine Helmer, Jean-Francois Dartigues, Helene Amieva, Michele Allard, Maxime Descoteaux, Gwenaelle Catheline, "Free water : a sensitive metric for age-related modifications of the cingulum white matter and its association with cognition" BioRxiv : 867606 (2019)
  - Theaud, Guillaume, Jean-Christophe Houde, Arnaud Boré, **Francois Rheault**, Felix Morency, and Maxime Descoteaux. "TractoFlow : A robust, efficient and reproducible diffusion MRI pipeline leveraging Nextflow & Singularity." (under review, Neuroimage NIMG-19-2346)
  - Mandonnet, Emmanuel, Vincent Marion, Antoni Valero-Cabré, Valentine Facque, Mélissa Dali, Marion Barberis, **Francois Rheault**, Emmanuelle Volle, Daniel S. Margulies, "Causal role of the Control network B in set-shifting during Trail Making Test part B : A multimodal analysis of a glioma surgery case" (under review, Cortex CORTEX-D-19-00360)
  - Vanderweyen, Davy, Guillaume Theaud, Jasmeen Sidhu, **Francois Rheault**, Silvio Sarubbo, Maxime Descoteaux, David Fortin, "The role of diffusion tractography in refining glial tumor resection" Brain Structure and Function (2020)

## Compétitions scientifiques, grand nombre de co-auteurs :

- Nath, Vishwesh, ..., **Francois Rheault**, ..., Bennett A. Landman "Tractography reproducibility challenge with empirical data (traced) : The 2017 ismrm diffusion study group challenge." Journal of Magnetic Resonance Imaging (2019).
- Schilling, Kurt G., ..., **Francois Rheault**, ..., Bennett A. Landman. "Limits to anatomical accuracy of diffusion tractography using modern approaches." NeuroImage 185 (2019) : 1-11.
- Maier-Hein, Klaus H, ..., **Francois Rheault**, ..., Maxime Descoteaux. "The challenge of mapping the human connectome based on diffusion tractography." Nature communications 8.1 (2017) : 1349.

## CONCLUSION

### Brevets et déclarations d'invention :

- “Outil automatisé d'identification de faisceaux de la matière blanche en tractographie cérébrale basée sur l'IRM de diffusion”, **Rheault, Francois**, Maxime Descoteaux et Jean-Christophe Houde, Imeka solutions inc. (Déclaration d'invention avec l'Université de Sherbrooke)

### Articles de Conférences

- **Rheault, Francois**, Etienne St-Onge, Jasmeen Sidhu, Quentin Chenot, Laurent Petit, and Maxime Descoteaux. "Bundle-specific tractography", Computational diffusion MRI workshop of Medical Image Computing and Computer Assisted Intervention (MICCAI 2018)

### Résumé de Conférences

- **Rheault, Francois**, Jean-Christophe Houde, Maxime Descoteaux, "Real time interaction with millions of streamlines", International Symposium in Magnetic Resonance in Medicine (ISMRM 2015, Toronto)
- **Rheault, Francois**, Jean-Christophe Houde, Nil Goyette, F Morency, Maxime Descoteaux, "MI-Brain, a software to handle tractograms and perform interactive virtual dissection", International Symposium in Magnetic Resonance in Medicine (ISMRM diffusion workshop 2016, Lisbonne)
- **Rheault, Francois**, Maxime Descoteaux, Louis Collins, "Streamlines non-linear registration using MR-Ultrasound for intra-operative brain shift correction", Organisation for Human Brain Mapping (OHBM 2016, Genève)
- Garyfallidis, Eleftherios, Marc-Alexandre Côté, **Francois Rheault**, Maxime Descoteaux, "QuickBundlesX : sequential clustering of millions of streamlines in multiple levels of detail at record execution time", International Symposium in Magnetic Resonance in Medicine (ISMRM 2016, Singapour)
- **Rheault, Francois**, Etienne St-Onge, Quentin Chenot, Laurent Petit, Maxime Descoteaux, "Bundle-specific tractography using voxel-wise orientation priors", International Symposium in Magnetic Resonance in Medicine (ISMRM 2018, Paris)
- Poulin, Philippe, **Francois Rheault**, Etienne St-Onge, Pierre-Marc Jodoin, Maxime Descoteaux, "Bundle-Wise Deep Tracker : Learning bundle-specific streamline paths", International Symposium in Magnetic Resonance in Medicine (ISMRM 2018, Paris)
- Chenot, Quentin, Nathalie Tzourio-Mazoyer, **Francois Rheault**, Maxime Descoteaux, and Laurent Petit, "Building a probabilistic atlas of the human corticospinal tract from 410 healthy participants by using enhanced bundle-specific tractography", International Symposium in Magnetic Resonance in Medicine (ISMRM 2018, Paris)
- Deslauriers, Samuel, Drew Parker, **Francois Rheault**, Rachid Deriche, Steven Brem, Maxime Descoteaux, Ragini Verma, "Edema-informed anatomically constrained particle filter tractography", Computational diffusion MRI workshop of Medical Image

## CONCLUSION

Computing and Computer Assisted Intervention (MICCAI 2018, Grenade)

- **Rheault, Francois**, Laurent Petit, and Maxime Descoteaux, "Tractostorm : Evaluation of intra and inter rater reproducibility in tractography dissection", International Symposium in Magnetic Resonance in Medicine (ISMRM 2019, Montréal)
- Sidhu, Jasmeen, **Francois Rheault**, and Maxime Descoteaux, "Needle in a Haystack : Finding connections of interest in the precentral gyrus from diffusion MRI based connectomics", International Symposium in Magnetic Resonance in Medicine (ISMRM 2019, Montréal)
- Ivanova Maria, **Francois Rheault**, and Nina Dronkers, "Finding critical language connections with multiple tractography algorithms : A new analytic approach", Organisation for Human Brain Mapping (OHBM 2020, Montréal)
- Chauvel Maelig, **Francois Rheault**, Cindy Rolland, Kevin Aubrain, Francois Leroy, Heloise de Vareilles, Gabriel Girard, Denis Riviere, Lucie Hertz-Pannier, Jean-Francois Mangin, Maxime Descoteaux and Jessica Dubois, "Exploring the micro-structural properties of the newborn sensori-motor network with diffusion MRI", Organisation for Human Brain Mapping (OHBM 2020, Montréal)

## Stages et collaborations hors-laboratoire

- **2015**, NatBrainLab, King's College London, Royaume-Uni, 3 mois, superviseur : Flavio Dell'Acqua
- **2015** NIST, Montreal Neurological Institute, Canada, 4 mois, superviseur : Louis Collins
- **2016** GIN-IMN, Université de Bordeaux, France, 3 mois, superviseur : Laurent Petit
- **2017** Département de neurochirurgie, Hôpital Lariboisière, France, 3 mois, superviseur : Emmanuel Mandonnet
- **2019** Knight Lab, Berkeley University, États-Unis, 3 mois, superviseuse : Maria Ivanova

“I think there is a world market for  
maybe five computers.”

---

— Thomas Watson, president of IBM (1943).

# Appendix A

## Annexes et schémas

### A.1 Tractographie spécifique aux faisceaux

Cette annexe contient les matériels supplémentaires de la publication présentée au Chapitre 5. Il s'agit d'éléments présents dans la publication tel que disponible en ligne, mais séparé dans le cadre de cette thèse pour faciliter la lecture.

#### A.1.1 BST, anatomical definition

The PT was extracted following the anatomical definition proposed by [Chenot et al. \[2018\]](#) based on a precise manual positioning of individual subcortical regions of interest (ROI) along the descending pathway of the PT. Three ROIs were therefore drawn in each hemisphere within the internal capsule, the midbrain and the medulla oblongata using both b0 and color FA maps. This was performed with TrackVis [[Wang et al., 2007](#)]. Additional anatomical ROIs were used from the JHU template [[Zhang et al., 2010](#)] once warped to the native diffusion space of each subject using ANTS non-linear registration. Streamlines passing through the 3 ROIs, terminating either in the frontal lobe or the parietal lobe, and not terminating in the cerebellum were considered to compose the PT.

The AF was extracted in each hemisphere from a first set of streamlines with one termination in the frontal lobe and the other either in the temporal lobe or in the occipital lobe. Then, streamlines that do not pass through the ventral part of the external capsule (vEC) were considered as composing the AF. Such streamlines are anatomically constraint to leave the frontal lobe from the superior part of the external capsule and then present

## A.1. TRACTOGRAPHIE SPÉCIFIQUE AUX FAISCEAUX

an arcuate shape defining the AF. The vEC ROI was composed of the region of the JHU template located in the ventral part of the external capsule, namely the inferior fronto-occipital and uncinate ROIs. Note that frontal, temporal and occipital lobes were composed of the frontal, temporal and occipital cortical gray matter and superficial white matter regions of the JHU template, respectively.

The CC was extracted from a first set of commissural streamlines with one termination in each hemisphere. A precentral-postcentral ROI, composed of the gray matter and superficial white matter part of the precentral and postcentral gyri of the JHU template was created for each hemisphere. Then, the present CC bundle was composed of callosal streamlines with one termination in each of the precentral-postcentral ROI and passing through the corpus callosum.

It is important to mention that streamlines considered for segmentation come from a tractogram generated with certain constraints and conditions. All streamlines are within the subject white matter, respect the length threshold (20mm-200mm), the angular threshold (45 degrees cone aperture for probabilistic and 20 degrees cone aperture for deterministic) and the minimum relative spherical function threshold of 0.1)

## A.1. TRACTOGRAPHIE SPÉCIFIQUE AUX FAISCEAUX

### A.1.2 BST, template creation overview

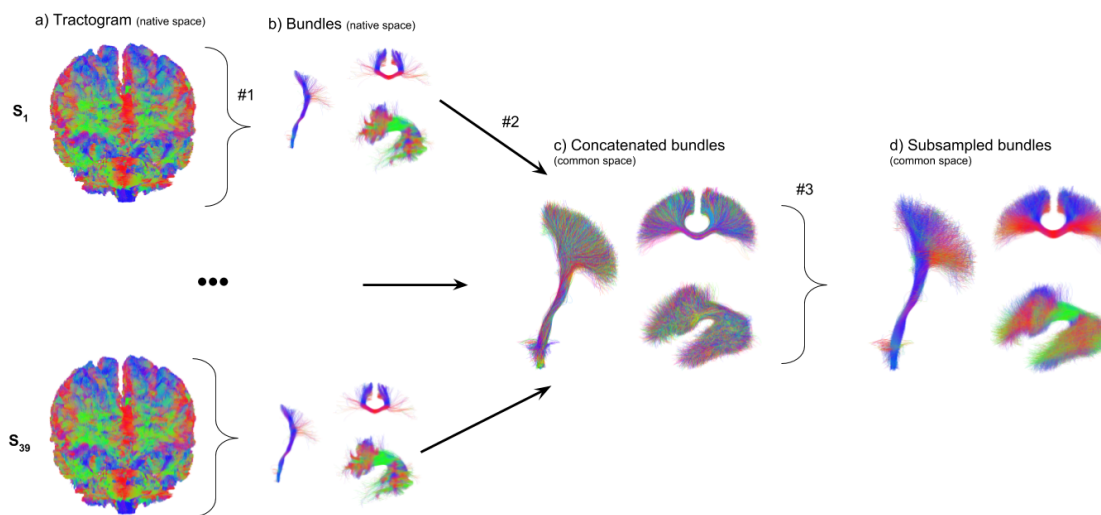


Figure A.1 – Overview of the template creation pipeline. From whole brain tractogram (a) bundles are extracted (b) and then warped and concatenated in a common space (c), finally the density bias is reduced by a spatial subsampling (d)

### A.1.3 BST, results from all tractography algorithms

In this section results from all tractography type are shown, comparaison to global tractography from the Medical Imaging Toolkit (MITK) [Kreher et al., 2008; Neher et al., 2012] was added. The tracking was performed with the original FOD and WM mask, default parameters and  $1 \times 10^8$  iterations. For a fair comparaison, each subject whole brain tractogram was filtered to be similar to the classic tractography. Tractogram was forced to end in the BOI adapted interface, subsampled to the same number of streamlines and then segmented using the same automatic method.

## A.1. TRACTOGRAPHIE SPÉCIFIQUE AUX FAISCEAUX

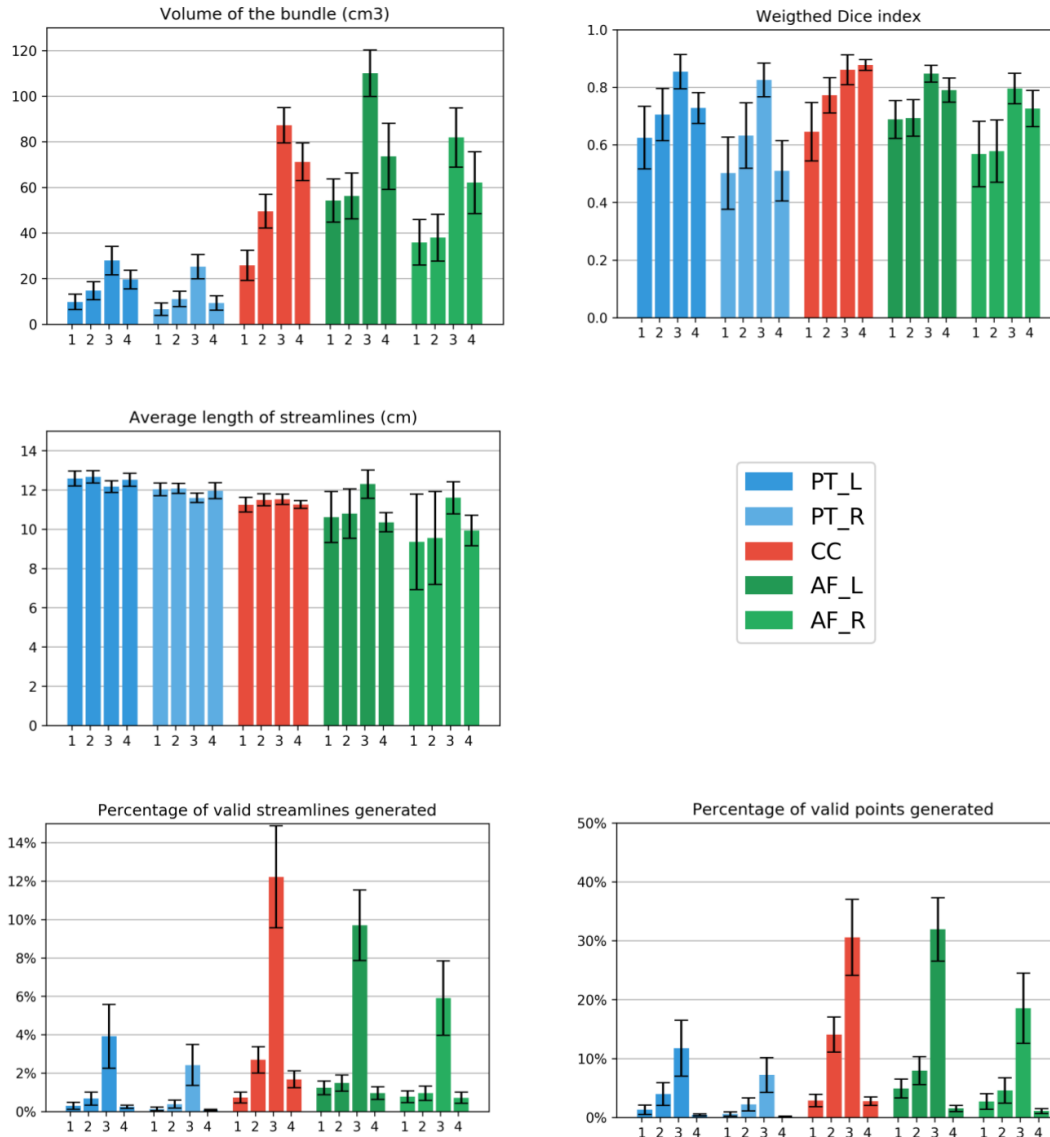


Figure A.2 – Results for each BOI across the 4 tracking variations with probabilistic particle filtering tractography, where the fourth method is the global tracking.



## A.1. TRACTOGRAPHIE SPÉCIFIQUE AUX FAISCEAUX

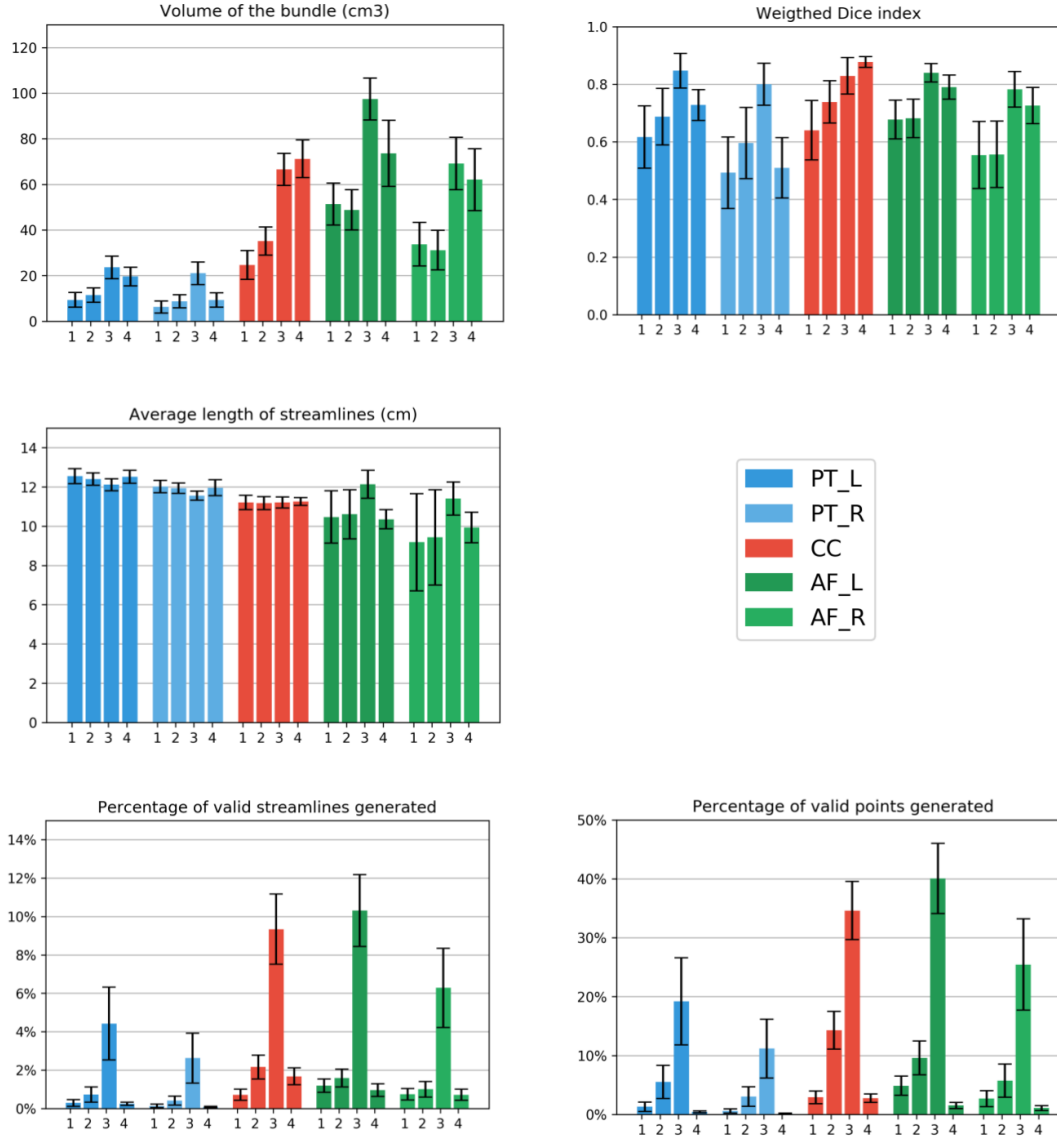


Figure A.3 – Results for each BOI across the 4 tracking variations with probabilistic tractography, where the fourth method is the global tracking.

## A.1. TRACTOGRAPHIE SPÉCIFIQUE AUX FAISCEAUX

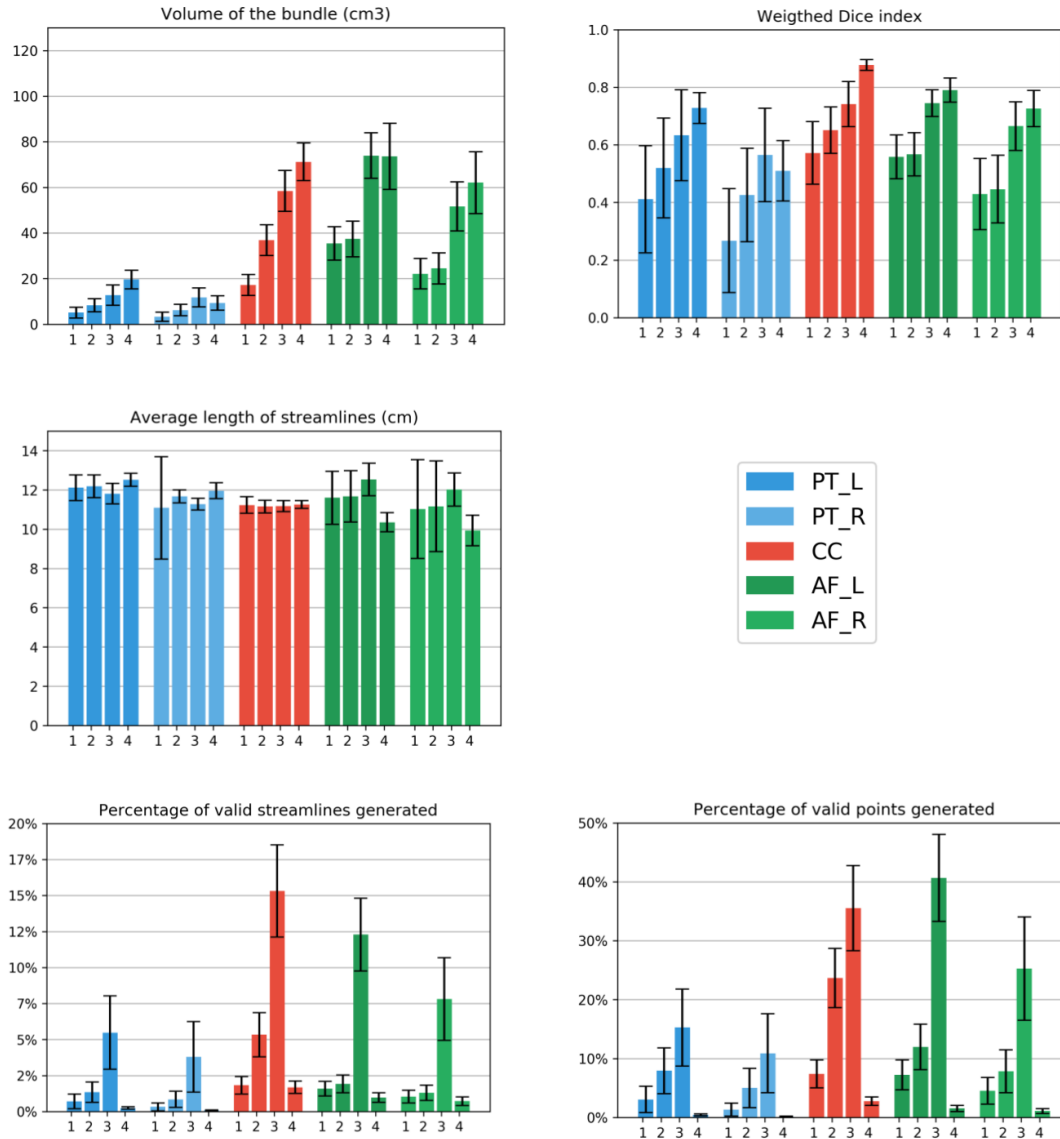


Figure A.4 – Results for each BOI across the 4 tracking variations with deterministic particle filtering tractography, where the fourth method is the global tracking.

## A.1. TRACTOGRAPHIE SPÉCIFIQUE AUX FAISCEAUX

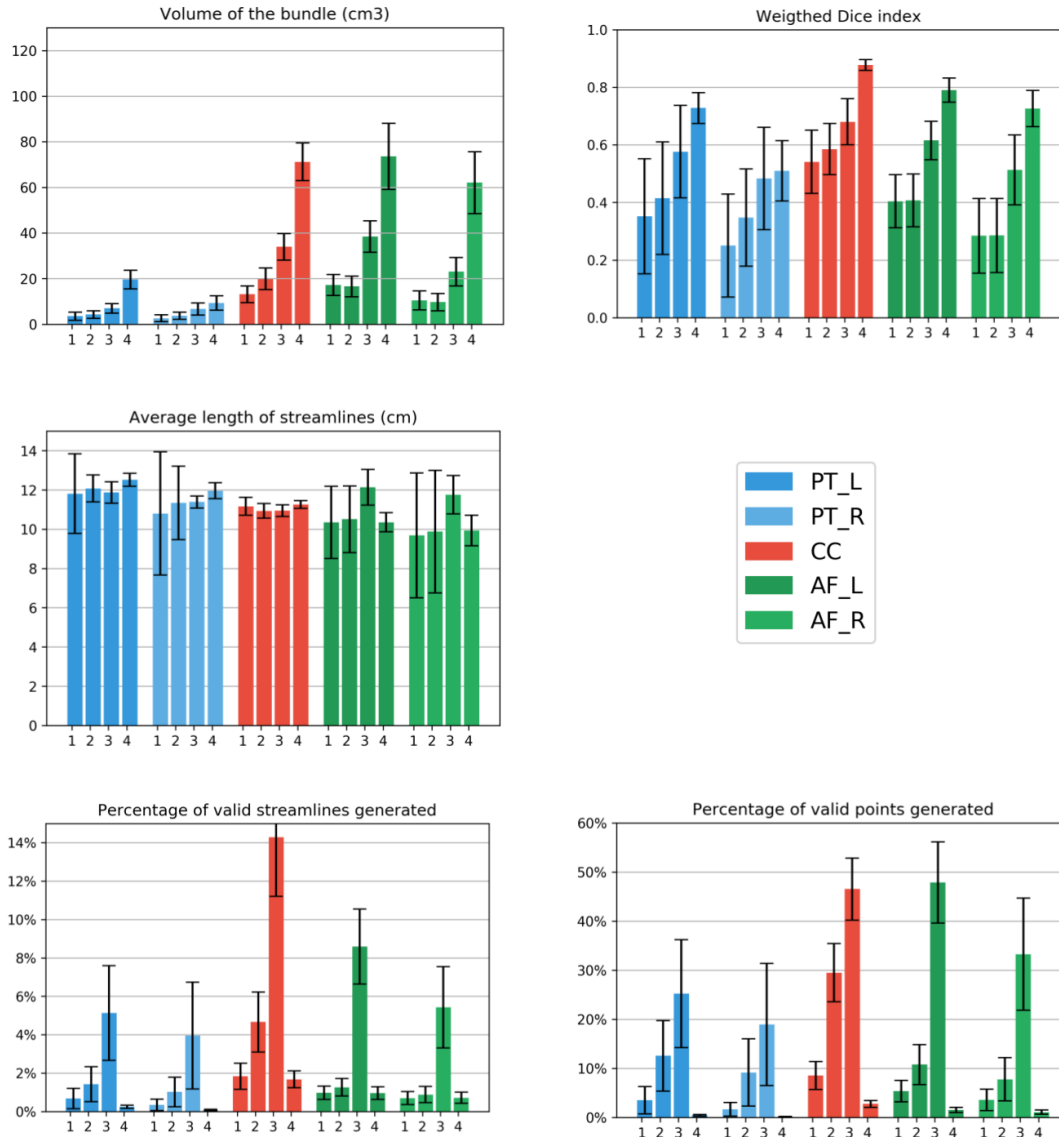


Figure A.5 – Results for each BOI across the 3 tracking variations with deterministic tractography, where the fourth method is the global tracking.

## A.2 Bundle-specific fornix reconstruction for dual-tracer PET-tractometry

### Résumé

La tractographie est connue pour ses limitations en ce qui concerne la reconstruction de faisceaux minces, ayant une courbure élevée et traversant beaucoup de volume partiel. Un de ces faisceaux est le fornix, un faisceau majeur ayant comme origine l'hippocampe, qui est particulièrement problématique pour la tractographie chez les participants âgés. L'atrophie de l'hippocampe et l'expansion ventriculaire rendent la reconstruction par tractographie difficile et parfois même impossible. Cette publication présente un algorithme de tractographie spécifique aux faisceaux visant à injecter des *a priori* de position, de forme et d'orientation pour reconstruire le fornix. Nous rapportons une augmentation du volume et une meilleure reproductibilité test-retest grâce à l'algorithme. Ces améliorations par rapport à la tractographie classique nous ont permis de quantifier le faisceau à l'aide de la tractométrie combinée à un traceur venant de tomographie par émission de position (TEP) chez des participants ayant un déficit cognitif léger (mild cognitive impairment, MCI). Des images multimodales de ces participants MCI après 6 mois de supplémentation kéto-génique quotidienne ont été acquises. Nous rapportons, pour la première fois, des différences significatives de mesures de diffusion et du traceur  $^{18}\text{F}$ -fluorodeoxyglucose (FDG) dans des sous-sections du fornix après l'utilisation de supplément céto-gène.

### Contributions de la publication

- Application de la méthode de tractographie spécifique aux faisceaux à un jeu de données de vieillissement.
- Quantification par tractométrie le long du fornix pour des mesures de diffusion et de TEP
- Observation de différences significatives grâce à la tractométrie entre les participants ayant eu des suppléments céto-gène et le groupe contrôle.

## A.2. BUNDLE-SPECIFIC FORNIX RECONSTRUCTION FOR DUAL-TRACER PET-TRACTOMETRY

### Contributions des auteurs

- Création de l’atlas de Fornix et reconstruction du faisceau par tractographie (Francois Rheault)
- Exécution du pipeline de tractométrie (Francois Rheault)
- Mise en place de l’étude, acquisition d’imagerie médicale des participants (Maggie Roy)
- Contrôle qualité, compilation et analyses des résultats (Maggie Roy)
- Écriture du texte et créations des figures et tables (Francois Rheault et Maggie Roy)
- Supervision du projet, correction et révision du manuscrit (Stephen Cunnane et Maxime Descoteaux)

### Commentaires

L’existence de cette publication est due au désir de détailler la méthodologie en profondeur. Une autre publication utilisant le même jeu de données est en production où plusieurs autres faisceaux sont analysés. Cette autre publication est plutôt axée sur l’application et les conséquences cliniques possibles, laissant de côté les détails méthodologiques de la tractographie. C’est pourquoi cette publication a été jugée nécessaire, mais séparée.

## A.2. BUNDLE-SPECIFIC FORNIX RECONSTRUCTION FOR DUAL-TRACER PET-TRACTOMETRY

Francois Rheault<sup>1</sup>, Maggie Roy<sup>1,2</sup>, Stephen Cunnane<sup>1,2</sup>, Maxime Descoteaux<sup>1</sup>,

<sup>1</sup>Sherbrooke Connectivity Imaging Laboratory (SCIL), Université de Sherbrooke, Canada

<sup>2</sup>Research Center on Aging, Université de Sherbrooke, Sherbrooke, Canada

### Abstract

Tractography is known to have problems reconstructing white matter bundles that are narrow, have high curvature, or go through partial volume voxels contaminated by CSF or gray matter. One such bundle is the fornix, the major output tract of the hippocampus, which is especially problematic with aging. Hippocampal atrophy and ventricular expansion make the fornix even harder (often impossible) to track with current state-of-the-art techniques. In this work, a bundle-specific tractography algorithm is proposed to fully reconstruct the fornix. By injecting shape, position, and orientation priors, fornix reconstruction is markedly improved. We report an increase in spatial coverage and better reproducibility across test-retest. These improvements over classical tractography algorithms also enable tractometry of the fornix to be combined with dual-tracer positron emission tomography (PET) data in participants with mild cognitive impairment (MCI). MCI participants underwent a multi-modal brain imaging before and after a 6-month daily ketogenic supplement. We report, for the first time, significant diffusion measures and 18F-fluorodeoxyglucose (FDG) uptake differences in specific sub-subsections of the fornix after the ketogenic supplement.

### A.2.1 Introduction

The fornix is a white matter (WM) bundle playing a vital role in memory and is therefore very relevant to analyze in cognitive decline associated with aging. The fornix originates from the hippocampi and join beneath the splenium of the corpus callosum to form the body of the fornix. The anterior body of the fornix then divides to enter both mamillary bodies. Reconstruction of the fornix white matter bundle for the tractography community is similar to hippocampus segmentation in the field of MRI segmentation [Kantarci, 2014], as it is an important tract of the memory structural network and thus, closely linked to neurodegenerative disease in aging population [Lister et Barnes, 2009].

In aged populations, hippocampal atrophy and ventricular expansion make the fornix extremely hard-to-track [Fletcher et al., 2013]. Specific acquisition schemes, such as FLAIR-DTI [Chou et al., 2005] which removes the signal from the ventricles, have been developed to facilitate tractography of the fornix. However, existing

## A.2. BUNDLE-SPECIFIC FORNIX RECONSTRUCTION FOR DUAL-TRACER PET-TRACTOMETRY

databases, such as Alzheimer’s Disease Neuroimaging Initiative (ADNI) [Petersen et al., 2010], often use classical diffusion MRI (dMRI) acquisitions tailored for diffusion tensor imaging (DTI) processing [Marek et al., 2011; Sudlow et al., 2015]. Alternatively, free-water (FW) elimination processing [Concha et al., 2005; Pasternak et al., 2009] can reduce CSF contamination causing partial volume effect around the ventricles and improve DTI tractography. Leading to an improved fornix reconstruction by typically increasing the spatial extent. However, FW-corrected DTI remains a single fiber modeling technique, which is known to be sub-optimal for tractography of bundles with high curvature and crossing fiber configurations [Girard et al., 2014; Tournier et al., 2008].

Without anatomical priors, researchers targeting difficult bundles of interest often use tractography with a brute force approach, which consists of an aggressive seeding strategy across the whole brain, often generating tens of millions of streamlines [Renauld et al., 2016], in the hope of having a few lucky streamlines surviving manual dissection [Catani et al., 2013] or automatic segmentation [Garyfallidis et al., 2018]. The low quality and difficulty of fornix tractography reconstruction explains why this bundle is rarely reported in the literature, particularly in the aging population.

In the present study, a bundle-specific approach was developed to reconstruct the fornix, increasing the spatial extent and improving the reproducibility across time points and participants. This allowed a dual-tracer PET fornix tractometry application study on 23 participants with mild cognitive impairment (MCI) at two time points, before and after a 6-month ketogenic supplement of MCT (medium chain triglyceride) or placebo.

## A.2.2 Methods

### Multi-modal diffusion & PET imaging

MCI participants were randomized in the placebo (N=11) or MCT (N=12) groups. Participants underwent multi-modal brain imaging before and after a 6-month daily supplement of 30 g/day MCT [Croteau et al., 2017]. The protocol consisted of an isotropic 1 mm T1-weighted (T1w) image, followed by a 1.8 mm isotropic high angular resolution diffusion imaging (HARDI; 60 directions,  $b=1500$  s/mm<sup>2</sup>), and a blip-up/blip-down  $b=0$  s/mm<sup>2</sup> acquisition to correct for distortions. Diffusion-weighted images were upsampled to 1 mm isotropic resolution, diffusion tensor measures, free-water index [Pasternak et al., 2009], and fiber orientation distribution function (fODF) [Descoteaux et al., 2009; Tournier et al., 2012] were computed using Dipy [Garyfallidis et al., 2014].

The MRI protocol was followed by a dual PET tracer session: AcAc (11C-acetoacetate) first, followed by FDG (18F-fluorodeoxyglucose) [Croteau et al., 2017]. SUV (standardized uptake values) summed-images from dynamic acquisitions were used. Finally, PET images were co-registered to the T1w image and then to upsampled dMRI data using ANTs [Avants et al., 2010]. Average PET tracer uptake and tract-profiling along 10 subsections of the fornix were measured. Data are expressed as delta ( $\Delta$ ) (post-supplementation minus pre-supplementation).

### Fornix reconstruction

The bundle-specific tractography (BST) approach is similar to the one proposed in [Rheault et al., 2018b] and is composed of three steps.

1. Build a template of streamlines that represents the shape and position of the fornix, the template must represented the full spatial extent of the fornix.
2. Build spatial anatomical priors that represents the endpoints and the general position of the bundle (seeding and tracking mask). To solve partial volume effect difficulties below the ventricles, tracking was allowed in GM to maximize the chance of fully reconstructing the fornix.
3. Build the orientation priors from the track-orientation distribution of the template of streamlines, and then create the enhanced fiber orientation distributions (E-fODF).

### Template creation

A population-specific template was created using *ANTs Multivariate Template Construction* [Avants et al., 2010] due to major structural changes appearing with age.



## A.2. BUNDLE-SPECIFIC FORNIX RECONSTRUCTION FOR DUAL-TRACER PET-TRACTOMETRY

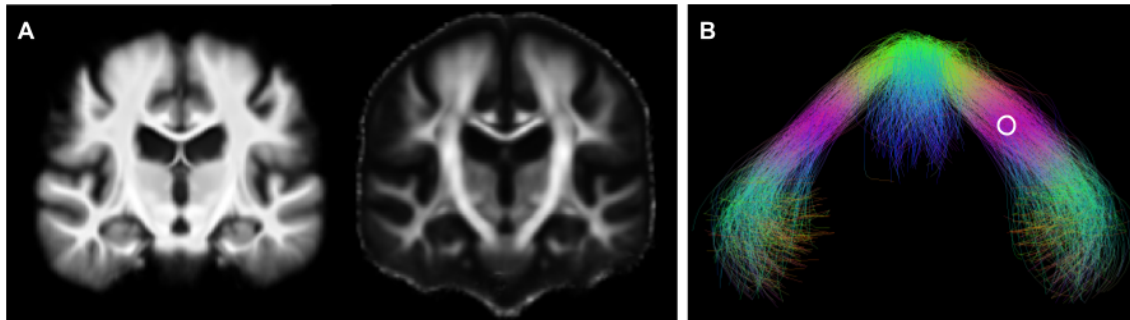


Figure A.6 – A) T1 and FA template from the 23 pre-supplementation acquisitions B) Representation of the average fornix for the 23 participants (zoomed). The white circle shows the region of origin of the fODF of Figure A.7.

T1w images and the fractional anisotropy (FA) maps from the pre-supplementation acquisitions ( $n=23$ ) were used as the input. The resulting average T1w and FA images can be seen in Figure A.6A. To create an initial streamline-template of the fornix, classical probabilistic tractography [Tournier et al., 2012], within WM and GM, was used on the pre-supplementation acquisitions. Each whole brain tractogram was then automatically segmented by moving streamlines to a common template space and then ROIs segmentation was used. Since each participant had a poorly saturated fornix, meaning most had less than 50 streamlines and that were not covering much of the expected volume, streamlines were concatenated in the common template space, forming a single, more dense, fornix template, as seen in Figure A.6B.

### Orientation enhancement & mask extraction

The newly created fornix template was then used to inform and enhance fiber orientation distribution function (fODF) needed by tractography, ultimately to help reconstruct the full fornix of each participant. First, the streamline-template was deformed to a participant native diffusion space. Then, at each voxel, an orientation prior was created from the streamline-template to reinforce the appropriate directions [Dhollander et al., 2014; Rheault et al., 2018b], as seen in Figure A.7. Seeding and tracking masks were also adapted to improve computational performance by restricting the regions where tractography is allowed [Renauld et al., 2016; Rheault et al., 2018b]. Tissue types were modified in a region under the ventricles particularly prone to partial volume effect, typically identified as CSF or GM, to allow tractography to propagate through the body of the fornix to the back of the thalamus.

## A.2. BUNDLE-SPECIFIC FORNIX RECONSTRUCTION FOR DUAL-TRACER PET-TRACTOMETRY

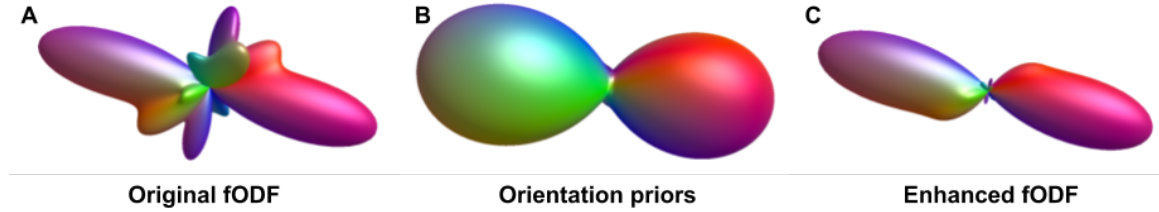


Figure A.7 – Application of the orientation prior in one voxel. The original fODF contains more than one fiber populations (A) and the orientation prior (B) reduces the influence of lobes that are not desirable or useful for the fornix reconstruction (C). The fODF shown is from the region within the white circle in Figure A.6B.

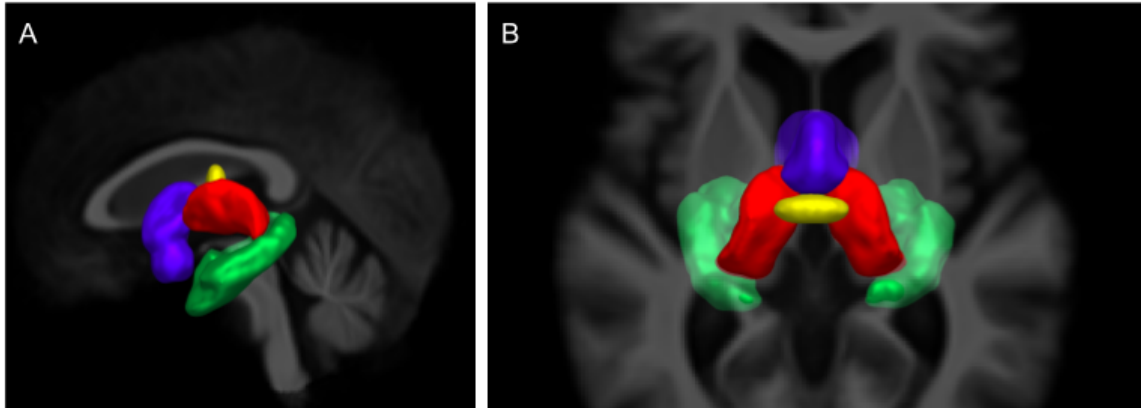


Figure A.8 – Sagittal view (A) and axial view (B) of the regions of interest used for the fornix segmentation in template space. Mammillary bodies in blue, body of the fornix in yellow, both thalami in red and both hippocampi in green.

## A.2. BUNDLE-SPECIFIC FORNIX RECONSTRUCTION FOR DUAL-TRACER PET-TRACTOMETRY

### Tractography

Anatomically-constrained particle filtering tractography (PFT) [Girard et al., 2014] with the bundle-specific probability maps and the bundle-specific enhanced-fODF was used to reconstruct the fornix (20 seeds per voxel, length threshold 40mm-160mm, and otherwise default parameters [Girard et al., 2014]). A classical approach was used as a comparison, i.e anatomically-constrained PFT with default parameters and no bundle-specific input (original fODF and probability WM-GM-CSF masks).

### Bundle segmentation

To compare the fornix across all datasets, an automatic segmentation was designed using anatomical landmarks in the population-specific template to reduce bundle segmentation variability. The thalamus and hippocampus were obtained with *Freesurfer* sub-cortical segmentation [Desikan et al., 2006] and mammillary bodies were manually delineated. To be considered valid, streamlines had to end in either the hippocampi or the mammillary bodies and had to pass through the body of the fornix [Catani et al., 2013]. The thalamus were used as an exclusion region of interest because the fornix has to curve around it. The anatomical regions of interest used for the segmentation can be seen in Figure A.8.

## A.2.3 Results

### Tractography

To evaluate fornix reconstructions, the spatial extent of the whole fornix was computed for each methods (classical vs enhanced), for each time point (pre vs post-supplementation) and each MCI group (placebo vs MCT). As seen in Figure A.9A, the classical method performs poorly in terms of volume against the enhanced version. Overall, the fornix reconstructed with the addition of priors were more dense and had a more coherent shape, as seen in Figure A.10.

Using the two time points as a test-retest datasets for each participant, overlap was measured using weighted Dice [Cousineau et al., 2017] to verify the degree of reproducibility of each fornix reconstruction. As seen in Figure A.9B, the classical method has severe difficulties to reconstruct the bundle across acquisitions, while the enhanced version obtains better overlap.

Overall, both the volume and weighted-Dice measures need to be taken into account together because the classical method obtains a low volume and low overlap, while the enhanced method obtains a higher volume and higher overlap, both which are needed for robust tractometry [Cousineau et al., 2017].

## A.2. BUNDLE-SPECIFIC FORNIX RECONSTRUCTION FOR DUAL-TRACER PET-TRACTOMETRY

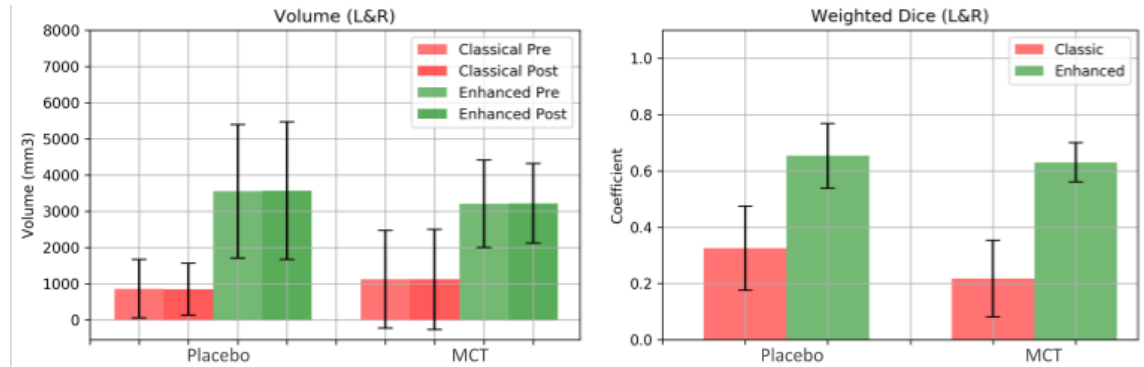


Figure A.9 – A) Average volume of the whole fornix comparison across acquisitions and groups for both methods (classical in red and the enhanced method in green). B) Weighted dice coefficient showing average overlap across time points for both groups. Data are expressed as mean and standard deviation.

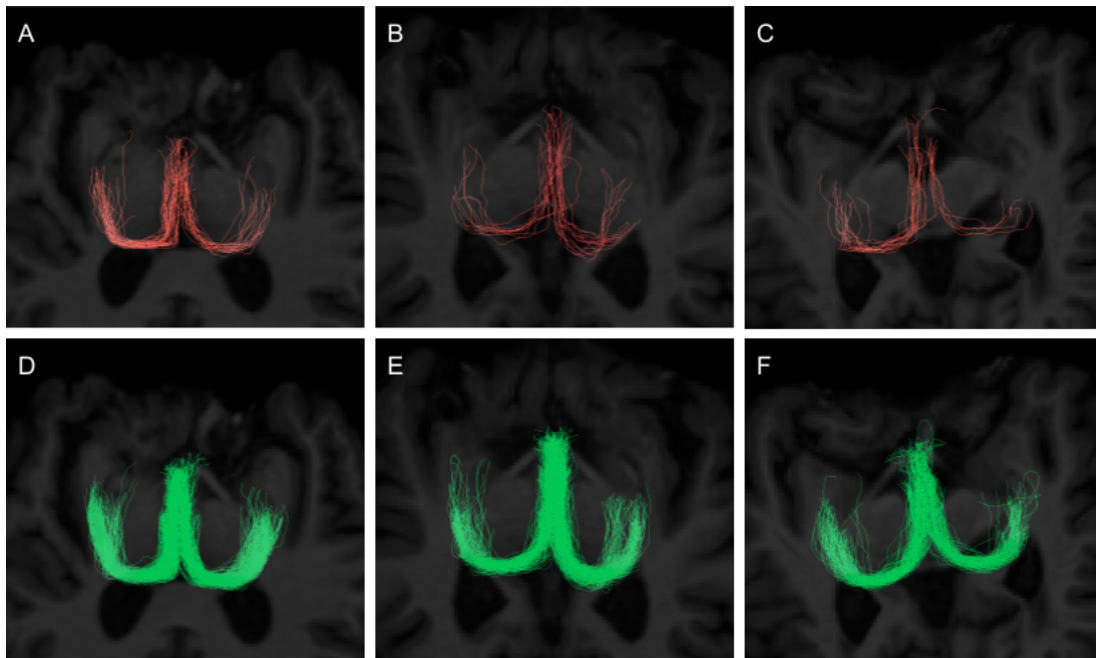


Figure A.10 – Comparison of the tractography output for 3 participants (pre-supplementation) between the classical method (A, B, C) and the enhanced method (D, E, F).

## A.2. BUNDLE-SPECIFIC FORNIX RECONSTRUCTION FOR DUAL-TRACER PET-TRACTOMETRY

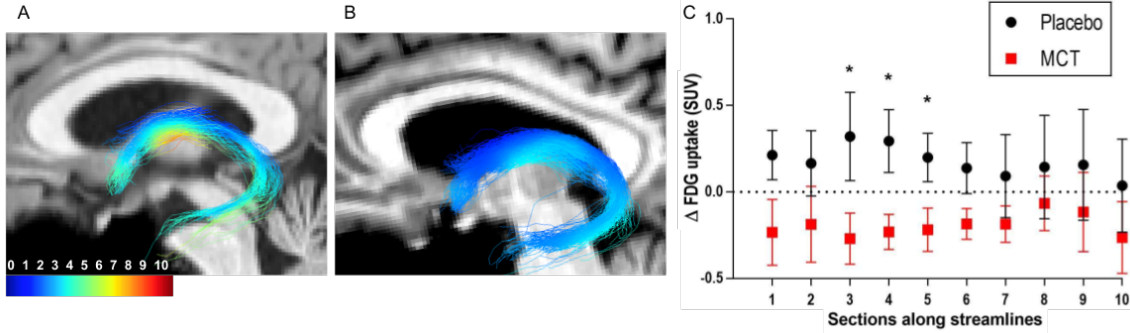


Figure A.11 – Sagittal view of the fornix from one MCI participant in the placebo group (A) and MCT group (B) at post-supplementation. The color scale represents 18F-fluorodeoxyglucose (FDG) uptake in SUV (standardized uptake values) along streamlines. The fornix was subsampled in 10 subsections: subsections 1-5 are located on the left part starting at the hippocampus; subsections 6-10 are located in the right part ending at the hippocampus. C) FDG uptake along the 10 subsubsections of the fornix. FDG uptake was significantly lower for the MCT group in subsections 3, 4 and 5. Data are expressed as delta (post- minus pre-supplementation) and *mean*  $\pm$  *SEM*.

### Tractometry

A preliminary tractometry analysis [Cousineau et al., 2017] was then performed on the enhanced fornix reconstructions using both the diffusion and dual-tracer PET data. Average PET tracer uptake and tract-profiling along 10 subsections (5 in the left part, 5 in the right part) of the fornix were measured, as seen in Figure A.11. Data are expressed as delta ( $\Delta$ ) (post minus pre-supplementation). All DTI measures (FA, mean, axial, and radial diffusivities), as well as free-water index were also submitted to the same tractometry analysis. First, lower  $\Delta$ FDG uptake was observed in the tract-profiles of the MCT group. subsections showing statistically significant differences are located in the left part of the body (subsection 3, 4 and 5;  $p = 0.05$ ,  $0.01$  and  $0.03$  respectively) as illustrated in Figure A.11. Tract-profiling of  $\Delta$ AcAc uptake showed no statistical difference with the current number of participants.

In Figure A.12, we show tract-profiles for  $\Delta$ free-water index, mean and radial diffusivity; higher values in the MCT group were found in the left part of the fornix (subsection 2 and 3) for the free-water and mean/radial diffusivities. When PET and diffusion data were averaged for the whole fornix, no statistically significant differences were observed post-intervention between the placebo and MCT groups.

## A.2. BUNDLE-SPECIFIC FORNIX RECONSTRUCTION FOR DUAL-TRACER PET-TRACTOMETRY

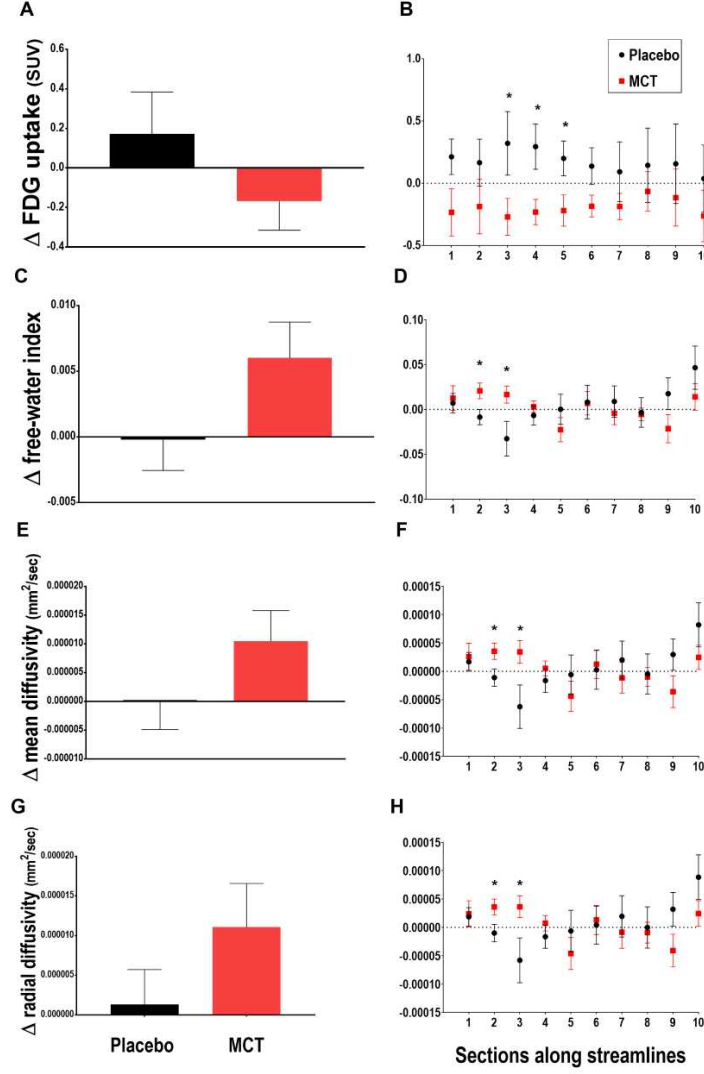


Figure A.12 – Average (left) and tract-profiles (right) for FDG PET and diffusion measures (free-water index, mean and radial diffusivities) in MCI participants of the placebo and MCT groups. The fornix was subsampled in 10 subsections: subsections 1-5 are located on the left part starting at the hippocampus; subsections 6-10 are located in the right part ending at the hippocampus. When PET and diffusion data were averaged for the whole fornix (A, C, E, G), no statistically significant difference were observed between the placebo and MCT groups, whereas subsections 2 and 3, on the left side of the fornix, show statistical significance for free-water index (D), mean (F) and radial diffusivities (H). Data are expressed as delta (post- minus pre-supplementation)

## A.2.4 Discussion

### Quality of the reconstruction

In this aged cohort, it is important to note that the fornix reconstruction with classical HARDI-based tractography methods was either impossible (complete fornix being un-trackable) or the quality so poor that a tractometry study was impossible. Hence, a dedicated bundle-specific fornix strategy was necessary.

In this work, quality of a tractography reconstruction is evaluated solely on spatial extent and reproducibility after a strict segmentation. This may not reflect the true anatomy of the fornix, as the full spatial extent remains unknown and was limited by our image resolution. On average, the test-retest Dice score was  $0.32 \pm 0.15$  for classical tractography and  $0.65 \pm 0.11$  for the proposed approach. Further improvements are needed to reach Dice scores of 0.8 or above. Seeding along the fornix instead of seeding from the endpoints could potentially produce an increased number of valid streamlines. Moreover, the initial choice of tracking direction could also be improved by selecting the largest fODF lobe, which is the one that represents the main fiber population of the fornix due to the orientation priors.

One of the major challenges for fornix-tractography is to obtain an adequate WM/GM/CSF segmentation. For this work, *FSL FAST* was used on the T1w images. On numerous occasions, partial volume effect due to enlarged ventricles was observed and WM was mis-classified as GM. More robust approaches using advanced tissues segmentation algorithms and priors suited for the aging brain are needed for anatomically constrained tractography.

It is also important to note that the creation of a population-specific template (T1, FA) was useful to facilitate the registration to a common space for analysis. Using a template such as MNI152, [Fonov et al., 2011] composed mainly of healthy participants, would have produced mediocre registration due to the atrophy of the cortex and enlarged ventricles in an older cohort [Avants et al., 2010].

### Alternative approaches to reconstruct the fornix

Other dMRI acquisitions or modeling could help tractography to accurately reconstruct the fornix. As mentioned earlier, FW elimination through modeling or through FLAIR-DTI can help reduce partial volume effect caused by the CSF around the ventricles and help to improve tractography [Chou et al., 2005]. However, the required acquisition is time consuming and most available databases do not use such dMRI sequences. Furthermore, the local model remains single-fiber based and comes with its own downside when other bundles of interest are targeted.

Another potential technique that could improve tractography of the fornix is local modeling approaches computed with multi-tissue response functions [Jeurissen et al.,



## A.2. BUNDLE-SPECIFIC FORNIX RECONSTRUCTION FOR DUAL-TRACER PET-TRACTOMETRY

2014]. This could reduce the impact of CSF contamination around the ventricles. This local modeling technique requires acquisitions with multiple b-values, but the work by [Dhollander et Connelly, 2016] recently proposed a multi-tissue local model that only requires single shell acquisition, which seems promising.

### Dual-tracer PET tractometry

Tract-profiling showed differences in FDG uptake between the two groups, which were not found when only the average FDG uptake in the fornix was compared. FDG uptake near the body and more left subsections, seemed more affected by the MCT supplementation compared to right subsections of the fornix. Therefore, the dense part of bundles may be more susceptible to changes in fuel (glucose) uptake. To our knowledge, this is the first report of FDG uptake in any WM fascicles in humans. The assessment of WM FDG uptake using PET enables the evaluation of WM energy metabolism *in vivo*. WM energy supply (axonal and oligodendrocytes), which serves mainly for resting potentials, myelin synthesis and intracellular trafficking of molecules, is crucial to sustain adequate axonal function [Harris et Attwell, 2012; Wender et al., 2000] and may be linked to the pathogenesis of MCI. Finally, it is critical to re-emphasize that this tractometry analysis would not have been possible without our enhanced bundle-specific fornix tracking. In most participants, the fornix was un-trackable or just composed of 10 streamlines or less in the mid-body of the fornix.

### Future work

Tractometry, using dual-tracer PET measures, along the fornix is now possible. Insightful measurements can be extracted using tractometry, which would have been missed by averaging of measures done on the full bundle. Future work will include a tractometry analysis using dual-tracer PET to assess whether a ketogenic supplement has impact on fuel uptake in other bundles, such as the posterior cingulum of MCI participants. Cerebral metabolic rate of FDG and AcAc will also be computed to compare quantitative measures between groups and this work will be extended to more participants and combined with grey matter analysis. A second objective is to map changes along the length of the fornix using various PET metric in healthy controls, MCI and Alzheimer’s patients to further evaluate the role of WM connections in cognitive impairment.



### **A.2.5 Conclusion**

Using a specific fornix streamline-template, position and orientation priors were injected to tractography in order to overcome fornix reconstruction difficulties in the aging brain. This process greatly improves the spatial coverage and reproducibility of tractography of the fornix in MCI. Without this tractography improvement, tract-based fornix investigation was simply impossible. Bundle-specific tractography could have a positive impact on aging studies using diffusion MRI to analyze specific WM fascicles.

“Failure is always an option.”

---

— Adam Savage, MythBusters.

# Bibliography

- Aganj, I., Lenglet, C., Sapiro, G., Yacoub, E., Ugurbil, K., Harel, N., 2010. Reconstruction of the orientation distribution function in single-and multiple-shell q-ball imaging within constant solid angle. *Magnetic resonance in medicine* 64, 554–566.
- Alexander-Bloch, A., Giedd, J.N., Bullmore, E., 2013. Imaging structural co-variance between human brain regions. *Nature Reviews Neuroscience* 14, 322.
- Alturkistani, H.A., Tashkandi, F.M., Mohammedsaleh, Z.M., 2016. Histological stains: a literature review and case study. *Global journal of health science* 8, 72.
- Apostolova, L.G., Zarow, C., Biado, K., Hurtz, S., Boccardi, M., Somme, J., Honarpisheh, H., Blanken, A.E., Brook, J., Tung, S., et al., 2015. Relationship between hippocampal atrophy and neuropathology markers: a 7T MRI validation study of the EADC-ADNI Harmonized Hippocampal Segmentation Protocol. *Alzheimer’s & Dementia* 11, 139–150.
- Atlason, H.E., Love, A., Sigurdsson, S., Gudnason, V., Ellingsen, L.M., 2019. SegAE: Un-supervised white matter lesion segmentation from brain MRIs using a CNN autoencoder. *NeuroImage: Clinical* 24, 102085.
- Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis* 12, 26–41.
- Avants, B.B., Tustison, N.J., Wu, J., Cook, P.A., Gee, J.C., 2011. An open source multivariate framework for n-tissue segmentation with evaluation on public data. *Neuroinformatics* 9, 381–400.
- Avants, B.B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., Gee,

## BIBLIOGRAPHY

- J.C., 2010. The optimal template effect in hippocampus studies of diseased populations. *Neuroimage* 49, 2457–2466.
- Axer, H., Beck, S., Axer, M., Schuchardt, F., Heepe, J., Flücken, A., Axer, M., Prescher, A., Witte, O.W., 2011a. Microstructural analysis of human white matter architecture using polarized light imaging: views from neuroanatomy. *Frontiers in neuroinformatics* 5, 28.
- Axer, H., Berks, G., Keyserlingk, D.G.V., 2000. Visualization of nerve fiber orientation in gross histological sections of the human brain. *Microscopy research and technique* 51, 481–492.
- Axer, M., Grassel, D., Kleiner, M., Dammers, J., Dickscheid, T., Reckfort, J., Hütz, T., Eiben, B., Pietrzyk, U., Zilles, K., et al., 2011b. High-resolution fiber tract reconstruction in the human brain by means of three-dimensional polarized light imaging. *Frontiers in neuroinformatics* 5, 34.
- Axer, M., Strohmer, S., Gräbel, D., Bucker, O., Dohmen, M., Reckfort, J., Zilles, K., Amunts, K., 2016. Estimating fiber orientation distribution functions in 3D-polarized light imaging. *Frontiers in neuroanatomy* 10, 40.
- Basser, P.J., Mattiello, J., LeBihan, D., 1994. Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance, Series B* 103, 247–254.
- Basser, P.J., Pajevic, S., Pierpaoli, C., Duda, J., Aldroubi, A., 2000. In vivo fiber tractography using DT-MRI data. *Magnetic resonance in medicine* 44, 625–632.
- Bayrak, R.G., Schilling, K.G., Greer, J.M., Hansen, C.B., Greer, C.M., Blaber, J.A., Williams, O., Beason-Held, L.L., Resnick, S.M., Rogers, B.P., et al., 2019. TractEM: Fast Protocols for Whole Brain Deterministic Tractography-Based White Matter Atlas. *bioRxiv* , 651935.
- Behrens, T.E., Berg, H.J., Jbabdi, S., Rushworth, M.F., Woolrich, M.W., 2007. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *Neuroimage* 34, 144–155.
- Behrens, T.E., Johansen-Berg, H., Woolrich, M., Smith, S., Wheeler-Kingshott, C., Boulby, P., Barker, G., Sillery, E., Sheehan, K., Ciccarelli, O., et al., 2003. Non-invasive mapping

## BIBLIOGRAPHY

- of connections between human thalamus and cortex using diffusion imaging. *Nature neuroscience* 6, 750.
- Benabid, A.L., 2003. Deep brain stimulation for Parkinson's disease. *Current opinion in neurobiology* 13, 696–706.
- Benedictis, A., Petit, L., Descoteaux, M., Marras, C.E., Barbareschi, M., Corsini, F., Dalabona, M., Chioffi, F., Sarubbo, S., 2016. New insights in the homotopic and heterotopic connectivity of the frontal portion of the human corpus callosum revealed by microdissection and diffusion tractography. *Human brain mapping* 37, 4718–4735.
- Bloch, F., 1946. Nuclear induction. *Physical review* 70, 460.
- Boccardi, M., Bocchetta, M., Apostolova, L.G., Barnes, J., Bartzokis, G., Corbetta, G., DeCarli, C., Firbank, M., Ganzola, R., Gerritsen, L., et al., 2015. Delphi definition of the EADC-ADNI Harmonized Protocol for hippocampal segmentation on magnetic resonance. *Alzheimer's & Dementia* 11, 126–138.
- Boccardi, M., Ganzola, R., Bocchetta, M., Pievani, M., Redolfi, A., Bartzokis, G., Camicioli, R., Csernansky, J.G., De Leon, M.J., deToledo Morrell, L., et al., 2011. Survey of protocols for the manual segmentation of the hippocampus: preparatory steps towards a joint EADC-ADNI harmonized protocol. *Journal of Alzheimer's disease* 26, 61–75.
- Bürgel, U., Mädler, B., Honey, C., Thron, A., Gilsbach, J., Coenen, V., 2009. Fiber tracking with distinct software tools results in a clear diversity in anatomical fiber tract portrayal. *Central European Neurosurgery-Zentralblatt für Neurochirurgie* 70, 27–35.
- Calamante, F., 2019. The Seven Deadly Sins of Measuring Brain Structural Connectivity Using Diffusion MRI Streamlines Fibre-Tracking. *Diagnostics* 9, 115.
- Calamante, F., Smith, R.E., Tournier, J.D., Raffelt, D., Connelly, A., 2015. Quantification of voxel-wise total fibre density: investigating the problems associated with track-count mapping. *Neuroimage* 117, 284–293.
- Canales-Rodríguez, E.J., Legarreta, J.H., Pizzolato, M., Renzonnet, G., Girard, G., Rafael-Patino, J., Barakovic, M., Romascano, D., Alemán-Gómez, Y., Radua, J., et al., 2019. Sparse wars: A survey and comparative study of spherical deconvolution algorithms for diffusion MRI. *NeuroImage* 184, 140–160.

## BIBLIOGRAPHY

- Catani, M., Allin, M.P., Husain, M., Pugliese, L., Mesulam, M.M., Murray, R.M., Jones, D.K., 2007. Symmetries in human brain language pathways correlate with verbal recall. *Proceedings of the National Academy of Sciences* 104, 17163–17168.
- Catani, M., De Schotten, M.T., 2008. A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *cortex* 44, 1105–1132.
- Catani, M., Dell’Acqua, F., De Schotten, M.T., 2013. A revised limbic system model for memory, emotion and behaviour. *Neuroscience & Biobehavioral Reviews* 37, 1724–1737.
- Catani, M., Howard, R.J., Pajevic, S., Jones, D.K., 2002. Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *Neuroimage* 17, 77–94.
- Caverzasi, E., Hervey-Jumper, S.L., Jordan, K.M., Lobach, I.V., Li, J., Panara, V., Racine, C.A., Sankaranarayanan, V., Amirbekian, B., Papinutto, N., et al., 2016. Identifying pre-operative language tracts and predicting postoperative functional recovery using HARDI q-ball fiber tractography in patients with gliomas. *Journal of neurosurgery* 125, 33–45.
- Ceotto, M., Tantardini, G.F., Aspuru-Guzik, A., 2011. Fighting the curse of dimensionality in first-principles semiclassical calculations: Non-local reference states for large number of dimensions. *The Journal of chemical physics* 135, 214108.
- Chamberland, M., Scherrer, B., Prabhu, S.P., Madsen, J., Fortin, D., Whittingstall, K., Descoteaux, M., Warfield, S.K., 2017. Active delineation of Meyer’s loop using oriented priors through MAGNETic tractography (MAGNET). *Human brain mapping* 38, 509–527.
- Chamberland, M., Whittingstall, K., Fortin, D., Mathieu, D., Descoteaux, M., 2014. Real-time multi-peak tractography for instantaneous connectivity display. *Frontiers in neuroinformatics* 8, 59.
- Chang, E.H., Argyelan, M., Aggarwal, M., Chandon, T.S.S., Karlsgodt, K.H., Mori, S., Malhotra, A.K., 2017. The role of myelination in measures of white matter integrity: combination of diffusion tensor imaging and two-photon microscopy of CLARITY intact brains. *Neuroimage* 147, 253–261.
- Chang, H.H., Zhuang, A.H., Valentino, D.J., Chu, W.C., 2009. Performance measure characterization for evaluating neuroimage segmentation algorithms. *Neuroimage* 47, 122–135.

## BIBLIOGRAPHY

- Chekir, A., Descoteaux, M., Garyfallidis, E., Côté, M.A., Boumghar, F.O., 2014. A hybrid approach for optimal automatic segmentation of White Matter tracts in HARDI, in: Biomedical Engineering and Sciences (IECBES), 2014 IEEE Conference on, IEEE. pp. 177–180.
- Cheng, J., Basser, P.J., 2018. Director Field Analysis (DFA): Exploring Local White Matter Geometric Structure in Diffusion MRI. *Medical image analysis* 43, 112–128.
- Chenot, Q., Tzourio-Mazoyer, N., Rheault, F., Descoteaux, M., Crivello, F., Zago, L., Mellet, E., Jobard, G., Joliot, M., Mazoyer, B., et al., 2018. A probabilistic atlas of the human pyramidal tract in 410 healthy participants. *bioRxiv* , 251108.
- Chenot, Q., Tzourio-Mazoyer, N., Rheault, F., Descoteaux, M., Crivello, F., Zago, L., Mellet, E., Jobard, G., Joliot, M., Mazoyer, B., et al., 2019. A population-based atlas of the human pyramidal tract in 410 healthy participants. *Brain Structure and Function* 224, 599–612.
- Chou, M.C., Lin, Y.R., Huang, T.Y., Wang, C.Y., Chung, H.W., Juan, C.J., Chen, C.Y., 2005. FLAIR diffusion-tensor MR tractography: comparison of fiber tracking with conventional imaging. *American journal of neuroradiology* 26, 591–597.
- Christiaens, D., Reisert, M., Dhollander, T., Sunaert, S., Suetens, P., Maes, F., 2015. Global tractography of multi-shell diffusion-weighted imaging data using a multi-tissue model. *Neuroimage* 123, 89–101.
- Chung, K., Deisseroth, K., 2013. CLARITY for mapping the nervous system. *Nature methods* 10, 508.
- Civier, O., Smith, R.E., Yeh, C.H., Connelly, A., Calamante, F., 2019. Is removal of weak connections necessary for graph-theoretical analysis of dense weighted structural connectomes from diffusion MRI? *NeuroImage* 194, 68–81.
- Collins, D.L., Pruessner, J.C., 2010. Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. *Neuroimage* 52, 1355–1366.
- Colon-Perez, L., Triplett, W., Bohsali, A., Corti, M., Nguyen, P., Patten, C., Mareci, T., Price, C., 2016. A majority rule approach for region-of-interest-guided streamline fiber tractography. *Brain imaging and behavior* 10, 1137–1147.

## BIBLIOGRAPHY

- Colon-Perez, L.M., Spindler, C., Goicochea, S., Triplett, W., Parekh, M., Montie, E., Carney, P.R., Price, C., Mareci, T.H., 2015. Dimensionless, scale invariant, edge weight metric for the study of complex structural networks. *PLOS one* 10, e0131493.
- Concha, L., Gross, D.W., Beaulieu, C., 2005. Diffusion tensor tractography of the limbic system. *American Journal of Neuroradiology* 26, 2267–2274.
- Côté, M.A., Girard, G., Boré, A., Garyfallidis, E., Houde, J.C., Descoteaux, M., 2013. Tractometer: towards validation of tractography pipelines. *Medical image analysis* 17, 844–857.
- Cousineau, M., Jodoin, P.M., Garyfallidis, E., Côté, M.A., Morency, F.C., Rozanski, V., Grand'Maison, M., Bedell, B.J., Descoteaux, M., 2017. A test-retest study on Parkinson's PPMI dataset yields statistically significant white matter fascicles. *NeuroImage: Clinical* 16, 222–233.
- Croteau, E., Castellano, C., Fortier, M., Bocti, C., Fulop, T., Paquet, N., Cunnane, S., 2017. A cross-sectional comparison of brain glucose and ketone metabolism in cognitively healthy older adults, mild cognitive impairment and early Alzheimer's disease. *Experimental gerontology* .
- Cunningham, D.J., 1906. *Textbook of anatomy*. W. Wood.
- Daducci, A., Dal Palú, A., Descoteaux, M., Thiran, J.P., 2016. Microstructure informed tractography: pitfalls and open challenges. *Frontiers in neuroscience* 10, 247.
- Daducci, A., Dal Palù, A., Lemkaddem, A., Thiran, J.P., 2015. COMMIT: convex optimization modeling for microstructure informed tractography. *IEEE transactions on medical imaging* 34, 246–257.
- Dayan, M., Kreutzer, S., Clark, C.A., 2015. Tractography of the optic radiation: a repeatability and reproducibility study. *NMR in Biomedicine* 28, 423–431.
- Dayan, M., Monohan, E., Pandya, S., Kuceyeski, A., Nguyen, T.D., Raj, A., Gauthier, S.A., 2016. Profilometry: a new statistical framework for the characterization of white matter pathways, with application to multiple sclerosis. *Human brain mapping* 37, 989–1004.
- De Benedictis, A., Nocerino, E., Menna, F., Remondino, F., Barbareschi, M., Rozzanigo, U., Corsini, F., Olivetti, E., Marras, C.E., Chioffi, F., et al., 2018. Photogrammetry



## BIBLIOGRAPHY

- of the human brain: A novel method for three-dimensional quantitative exploration of the structural connectivity in neurosurgery and neurosciences. *World Neurosurgery* 115, e279–e291.
- De Erausquin, G.A., Alba-Ferrara, L., 2013. What does anisotropy measure? Insights from increased and decreased anisotropy in selective fiber tracts in schizophrenia. *Frontiers in integrative neuroscience* 7, 9.
- Dejerine, J., Dejerine-Klumpke, A., 1895. *Anatomie des centres nerveux: Méthodes générales d'étude-embryologie-histogénèse et histologie. Anatomie du cerveau. volume 1.* Rueff.
- Déjerine, J., Dejerine-Klumpke, A., 1901. *Anatomie des centres nerveux, Tome II.* Paris: Rueff et Cie. éditeurs .
- Dell'Acqua, F., Scifo, P., Rizzo, G., Catani, M., Simmons, A., Scotti, G., Fazio, F., 2010. A modified damped Richardson–Lucy algorithm to reduce isotropic background effects in spherical deconvolution. *Neuroimage* 49, 1446–1458.
- Dell'Acqua, F., Simmons, A., Williams, S.C., Catani, M., 2013. Can spherical deconvolution provide more information than fiber orientations? Hindrance modulated orientational anisotropy, a true-tract specific index to characterize white matter diffusion. *Human brain mapping* 34, 2464–2483.
- Dell'Acqua, F., Tournier, J.D., 2019. Modelling white matter with spherical deconvolution: How and why? *NMR in Biomedicine* 32, e3945.
- Descoteaux, M., 2015. High angular resolution diffusion imaging (hardi). *Wiley Encyclopedia of Electrical and Electronics Engineering* .
- Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R., 2007. Regularized, fast, and robust analytical Q-ball imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 58, 497–510.
- Descoteaux, M., Deriche, R., Knosche, T.R., Anwander, A., 2009. Deterministic and probabilistic tractography based on complex fibre orientation distributions. *IEEE transactions on medical imaging* 28, 269–286.
- Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., et al., 2006. An automated labeling

## BIBLIOGRAPHY

- system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980.
- Dhollander, T., Connelly, A., 2016. A novel iterative approach to reap the benefits of multi-tissue CSD from just single-shell ( $b=0$ ) diffusion MRI data, in: *Proc ISMRM*, p. 3010.
- Dhollander, T., Emsell, L., Van Hecke, W., Maes, F., Sunaert, S., Suetens, P., 2014. Track orientation density imaging (TODI) and track orientation distribution (TOD) based tractography. *NeuroImage* 94, 312–336.
- Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E., Notredame, C., 2017. Nextflow enables reproducible computational workflows. *Nature biotechnology* 35, 316–319.
- Dice, L.R., 1945. Measures of the amount of ecologic association between species. *Ecology* 26, 297–302.
- Donahue, C.J., Sotiropoulos, S.N., Jbabdi, S., Hernandez-Fernandez, M., Behrens, T.E., Dyrby, T.B., Coalson, T., Kennedy, H., Knoblauch, K., Van Essen, D.C., et al., 2016. Using diffusion tractography to predict cortical connection strength and distance: a quantitative comparison with tracers in the monkey. *Journal of Neuroscience* 36, 6758–6770.
- Ebeling, U., Reulen, H.J., 1992. Subcortical topography and proportions of the pyramidal tract. *Acta neurochirurgica* 118, 164–171.
- Eickhoff, S.B., Stephan, K.E., Mohlberg, H., Grefkes, C., Fink, G.R., Amunts, K., Zilles, K., 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25, 1325–1335.
- Entis, J.J., Doerga, P., Barrett, L.F., Dickerson, B.C., 2012. A reliable protocol for the manual segmentation of the human amygdala and its subregions using ultra-high resolution MRI. *Neuroimage* 60, 1226–1235.
- Esmail-Zadeh, M., Soltanian-Zadeh, H., Jafari-Khouzani, K., 2010. SPHARM-based shape analysis of hippocampus for lateralization in mesial temporal lobe epilepsy, in: *Electrical Engineering (ICEE), 2010 18th Iranian Conference on, IEEE*. pp. 39–44.

## BIBLIOGRAPHY

- Essayed, W.I., Zhang, F., Unadkat, P., Cosgrove, G.R., Golby, A.J., O'Donnell, L.J., 2017. White matter tractography for neurosurgical planning: A topography-based review of the current state of the art. *NeuroImage: Clinical* 15, 659–672.
- Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A.R., et al., 2016. The human brainnetome atlas: a new brain atlas based on connectional architecture. *Cerebral cortex* 26, 3508–3526.
- Fernandes, H.M., Van Hartevelt, T.J., Boccard, S.G., Owen, S.L., Cabral, J., Deco, G., Green, A.L., Fitzgerald, J.J., Aziz, T.Z., Kringelbach, M.L., 2015. Novel fingerprinting method characterises the necessary and sufficient structural connectivity from deep brain stimulation electrodes for a successful outcome. *New Journal of Physics* 17, 015001.
- Fernández-Miranda, J.C., Rhoton Jr, A.L., Álvarez-Linera, J., Kakizawa, Y., Choi, C., de Oliveira, E.P., 2008. Three-dimensional microsurgical and tractographic anatomy of the white matter of the human brain. *Neurosurgery* 62, SHC989–SHC1028.
- Fernández-Miranda, J.C., Wang, Y., Pathak, S., Stefaneau, L., Verstynen, T., Yeh, F.C., 2015. Asymmetry, connectivity, and segmentation of the arcuate fascicle in the human brain. *Brain Structure and Function* 220, 1665–1680.
- Filatova, O.G., van Vliet, L.J., Schouten, A.C., Kwakkel, G., van der Helm, F.C., Vos, F.M., 2018. Comparison of multi-tensor diffusion models' performance for white matter integrity estimation in chronic stroke. *Frontiers in neuroscience* 12, 247.
- Fillard, P., Poupon, C., Mangin, J.F., 2009. A novel global tractography algorithm based on an adaptive spin glass model, in: *International conference on medical image computing and computer-assisted intervention*, Springer. pp. 927–934.
- Fletcher, E., Raman, M., Huebner, P., Liu, A., Mungas, D., Carmichael, O., DeCarli, C., 2013. Loss of fornix white matter volume as a predictor of cognitive impairment in cognitively normal elderly individuals. *JAMA neurology* 70, 1389–1395.
- Fonov, V., Evans, A.C., Botteron, K., Almli, C.R., McKinstry, R.C., Collins, D.L., Group, B.D.C., et al., 2011. Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage* 54, 313–327.

## BIBLIOGRAPHY

- Forkel, S.J., de Schotten, M.T., Kawadler, J.M., Dell’Acqua, F., Danek, A., Catani, M., 2014. The anatomy of fronto-occipital connections from early blunt dissections to contemporary tractography. *Cortex* 56, 73–84.
- Frisoni, G.B., Jack Jr, C.R., Bocchetta, M., Bauer, C., Frederiksen, K.S., Liu, Y., Preboske, G., Swihart, T., Blair, M., Cavedo, E., et al., 2015. The EADC-ADNI Harmonized Protocol for manual hippocampal segmentation on magnetic resonance: evidence of validity. *Alzheimer’s & Dementia* 11, 111–125.
- Friston, K.J., Glaser, D.E., Henson, R.N., Kiebel, S., Phillips, C., Ashburner, J., 2002a. Classical and Bayesian inference in neuroimaging: applications. *Neuroimage* 16, 484–512.
- Friston, K.J., Penny, W., Phillips, C., Kiebel, S., Hinton, G., Ashburner, J., 2002b. Classical and Bayesian inference in neuroimaging: theory. *NeuroImage* 16, 465–483.
- Garyfallidis, E., 2013. Towards an accurate brain tractography. Ph.D. thesis. University of Cambridge.
- Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., Van Der Walt, S., Descoteaux, M., Nimmo-Smith, I., 2014. Dipy, a library for the analysis of diffusion MRI data. *Frontiers in neuroinformatics* 8, 8.
- Garyfallidis, E., Brett, M., Correia, M.M., Williams, G.B., Nimmo-Smith, I., 2012. Quickbundles, a method for tractography simplification. *Frontiers in neuroscience* 6, 175.
- Garyfallidis, E., Brett, M., Nimmo-Smith, I., 2010. Fast dimensionality reduction for brain tractography clustering, in: 16th Annual Meeting of the Organization for Human Brain Mapping.
- Garyfallidis, E., Côté, M.A., Rheault, F., Descoteaux, M., 2016. QuickBundlesX: sequential clustering of millions of streamlines in multiple levels of detail at record execution time. *ISMRM2016 (Singapore)* .
- Garyfallidis, E., Côté, M.A., Rheault, F., Sidhu, J., Hau, J., Petit, L., Fortin, D., Cunanne, S., Descoteaux, M., 2018. Recognition of white matter bundles using local and global streamline-based registration and clustering. *NeuroImage* 170, 283–295.

## BIBLIOGRAPHY

- Garyfallidis, E., Ocegueda, O., Wassermann, D., Descoteaux, M., 2015. Robust and efficient linear registration of white-matter fascicles in the space of streamlines. *NeuroImage* 117, 124–140.
- Gauvin, A., 2016. Assurance qualité en dissection virtuelle des faisceaux de la matière blanche par tractographie. Master’s thesis. Université de Sherbrooke.
- Ghaziri, J., Tucholka, A., Girard, G., Boucher, O., Houde, J.C., Descoteaux, M., Obaid, S., Gilbert, G., Rouleau, I., Nguyen, D.K., 2018. Subcortical structural connectivity of insular subregions. *Scientific reports* 8, 8596.
- Ghaziri, J., Tucholka, A., Girard, G., Houde, J.C., Boucher, O., Gilbert, G., Descoteaux, M., Lippé, S., Rainville, P., Nguyen, D.K., 2015. The corticocortical structural connectivity of the human insula. *Cerebral cortex* 27, 1216–1228.
- Girard, G., Fick, R., Descoteaux, M., Deriche, R., Wassermann, D., 2015. AxTract: microstructure-driven tractography based on the ensemble average propagator, in: *International Conference on Information Processing in Medical Imaging*, Springer. pp. 675–686.
- Girard, G., Whittingstall, K., Deriche, R., Descoteaux, M., 2014. Towards quantitative connectivity analysis: reducing tractography biases. *Neuroimage* 98, 266–278.
- Gisev, N., Bell, J.S., Chen, T.F., 2013. Interrater agreement and interrater reliability: key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy* 9, 330–338.
- Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., et al., 2016. A multi-modal parcellation of human cerebral cortex. *Nature* 536, 171.
- Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., et al., 2013. The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* 80, 105–124.
- Glozman, T., Bruckert, L., Pestilli, F., Yecies, D.W., Guibas, L.J., Yeom, K.W., 2018. Framework for shape analysis of white matter fiber bundles. *NeuroImage* 167, 466–477.

## BIBLIOGRAPHY

- Griffa, A., Baumann, P.S., Thiran, J.P., Hagmann, P., 2013. Structural connectomics in brain diseases. *Neuroimage* 80, 515–526.
- Griffiths, K., Grieve, S., Kohn, M., Clarke, S., Williams, L., Korgaonkar, M., 2016. Altered gray matter organization in children and adolescents with ADHD: a structural covariance connectome study. *Translational psychiatry* 6, e947.
- Groeschel, S., Tournier, J.D., Northam, G.B., Baldeweg, T., Wyatt, J., Vollmer, B., Connelly, A., 2014. Identification and interpretation of microstructural abnormalities in motor pathways in adolescents born preterm. *NeuroImage* 87, 209–219.
- Gryska, E.A., Schneiderman, J., Heckemann, R.A., 2019. Automatic brain lesion segmentation on standard MRIs of the human head: a scoping review protocol. *BMJ open* 9, e024824.
- Guevara, P., Duclap, D., Poupon, C., Marrakchi-Kacem, L., Fillard, P., Le Bihan, D., Leboyer, M., Houenou, J., Mangin, J.F., 2012. Automatic fiber bundle segmentation in massive tractography datasets using a multi-subject bundle atlas. *Neuroimage* 61, 1083–1099.
- Guevara, P., Poupon, C., Rivière, D., Cointepas, Y., Descoteaux, M., Thirion, B., Mangin, J.F., 2011. Robust clustering of massive tractography datasets. *NeuroImage* 54, 1975–1993.
- Gwet, K.L., 2012. Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among multiple raters. Advanced Analytics, LLC .
- Hagmann, P., 2005. From diffusion MRI to brain connectomics. Technical Report. EPFL.
- Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Wedeen, V.J., Sporns, O., 2008. Mapping the structural core of human cerebral cortex. *PLoS biology* 6, e159.
- Harris, J.J., Attwell, D., 2012. The energetics of CNS white matter. *Journal of Neuroscience* 32, 356–371.
- Hau, J., Sarubbo, S., Houde, J.C., Corsini, F., Girard, G., Deledalle, C., Crivello, F., Zago, L., Mellet, E., Jobard, G., et al., 2017. Revisiting the human uncinate fasciculus, its subcomponents and asymmetries with stem-based tractography and microdissection validation. *Brain Structure and Function* 222, 1645–1662.

## BIBLIOGRAPHY

- Hau, J., Sarubbo, S., Perchey, G., Crivello, F., Zago, L., Mellet, E., Jobard, G., Joliot, M., Mazoyer, B.M., Tzourio-Mazoyer, N., et al., 2016. Cortical terminations of the inferior fronto-occipital and uncinate fasciculi: anatomical stem-based virtual dissection. *Frontiers in neuroanatomy* 10, 58.
- Hyde, J.S., Jesmanowicz, A., 2012. Cross-correlation: an fMRI signal-processing strategy. *NeuroImage* 62, 848–851.
- Iglesias, J.E., Sabuncu, M.R., 2015. Multi-atlas segmentation of biomedical images: a survey. *Medical image analysis* 24, 205–219.
- Ingahalikar, M., Smith, A., Parker, D., Satterthwaite, T.D., Elliott, M.A., Ruparel, K., Hakonarson, H., Gur, R.E., Gur, R.C., Verma, R., 2014. Sex differences in the structural connectome of the human brain. *Proceedings of the National Academy of Sciences* 111, 823–828.
- Jbabdi, S., Johansen-Berg, H., 2011. Tractography: where do we go from here? *Brain connectivity* 1, 169–183.
- Jeurissen, B., Descoteaux, M., Mori, S., Leemans, A., 2017. Diffusion MRI fiber tractography of the brain. *NMR in Biomedicine* .
- Jeurissen, B., Leemans, A., Tournier, J.D., Jones, D.K., Sijbers, J., 2013. Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. *Human brain mapping* 34, 2747–2766.
- Jeurissen, B., Tournier, J.D., Dhollander, T., Connelly, A., Sijbers, J., 2014. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *NeuroImage* 103, 411–426.
- Jiang, H., Van Zijl, P.C., Kim, J., Pearlson, G.D., Mori, S., 2006. DtiStudio: resource program for diffusion tensor computation and fiber bundle tracking. *Computer methods and programs in biomedicine* 81, 106–116.
- Johansen-Berg, H., Behrens, T., Robson, M., Drobnjak, I., Rushworth, M., Brady, J., Smith, S., Higham, D., Matthews, P., 2004. Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proceedings of the National Academy of Sciences* 101, 13335–13340.

## BIBLIOGRAPHY

- Johansen-Berg, H., Behrens, T.E., 2013. Diffusion MRI: from quantitative measurement to in vivo neuroanatomy. Academic Press.
- Jones, D., Simmons, A., Williams, S., Horsfield, M., 1998. Non-invasive assessment of structural connectivity in white matter by diffusion tensor MRI, in: Book of Abstracts: Sixth Annual Meeting of the International Society for Magnetic Resonance in Medicine, Int. Soc. Magn. Reson. Med Berkeley. p. 531.
- Jones, D.K., 2010. Challenges and limitations of quantifying brain connectivity in vivo with diffusion MRI. *Imaging in Medicine* 2, 341.
- Jones, D.K., Cercignani, M., 2010. Twenty-five pitfalls in the analysis of diffusion MRI data. *NMR in Biomedicine* 23, 803–820.
- Jones, D.K., Knösche, T.R., Turner, R., 2013. White matter integrity, fiber count, and other fallacies: the do’s and don’ts of diffusion MRI. *Neuroimage* 73, 239–254.
- Kantarci, K., 2014. Fractional anisotropy of the fornix and hippocampal atrophy in Alzheimer’s disease. *Frontiers in aging neuroscience* 6, 316.
- Kaur, S., Powell, S., He, L., Pierson, C.R., Parikh, N.A., 2014. Reliability and repeatability of quantitative tractography methods for mapping structural white matter connectivity in preterm and term infants at term-equivalent age. *PLoS One* 9, e85807.
- Kimura-Ohba, S., Yang, Y., Thompson, J., Kimura, T., Salayandia, V.M., Cosse, M., Yang, Y., Sillerud, L.O., Rosenberg, G.A., 2016. Transient increase of fractional anisotropy in reversible vasogenic edema. *Journal of Cerebral Blood Flow & Metabolism* 36, 1731–1743.
- Kleesiek, J., Petersen, J., Döring, M., Maier-Hein, K., Köthe, U., Wick, W., Hamprecht, F.A., Bendszus, M., Biller, A., 2016. Virtual raters for reproducible and objective assessments in radiology. *Scientific reports* 6, 25007.
- Kreher, B., Mader, I., Kiselev, V., 2008. Gibbs tracking: a novel approach for the reconstruction of neuronal pathways. *Magnetic Resonance in Medicine* 60, 953–963.
- Kreher, B., Schneider, J., Mader, I., Martin, E., Hennig, J., Il’Yasov, K., 2005. Multitensor approach for analysis and tracking of complex fiber configurations. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 54, 1216–1225.



## BIBLIOGRAPHY

- Kreilkamp, B.A., Lisanti, L., Glenn, G.R., Wieshmann, U.C., Das, K., Marson, A.G., Keller, S.S., 2019. Comparison of Manual and Automated Fiber Quantification Tractography in Patients with Temporal Lobe Epilepsy. *NeuroImage: Clinical* , 102024.
- Kristensson, K., Olsson, Y., 1973. Diffusion pathways and retrograde axonal transport of protein tracers in peripheral nerves. *Progress in neurobiology* 1, 85–109.
- Kurtzer, G.M., Sochat, V., Bauer, M.W., 2017. Singularity: Scientific containers for mobility of compute. *PloS one* 12, e0177459.
- Lampert, T.A., Stumpf, A., Gañarski, P., 2016. An empirical study into annotator agreement, ground truth estimation, and algorithm evaluation. *IEEE Transactions on Image Processing* 25, 2557–2572.
- Langerak, T.R., van der Heide, U.A., Kotte, A.N., Berendsen, F.F., Pluim, J.P., 2015. Improving label fusion in multi-atlas based segmentation by locally combining atlas selection and performance estimation. *Computer Vision and Image Understanding* 130, 71–79.
- Lauterbur, P.C., 1973. Image formation by induced local interactions: examples employing nuclear magnetic resonance. *nature* 242, 190–191.
- Lavail, J.H., Lavail, M.M., 1972. Retrograde axonal transport in the central nervous system. *Science* 176, 1416–1417.
- Le Bihan, D., Breton, E., 1985. Imagerie de diffusion in-vivo par résonance magnétique nucléaire. *Comptes-Rendus de l’Académie des Sciences* 93, 27–34.
- Lee Masson, H., Wallraven, C., Petit, L., 2017. “Can touch this”: Cross-modal shape categorization performance is associated with microstructural characteristics of white matter association pathways. *Human brain mapping* 38, 842–854.
- Leemans, A., Jeurissen, B., Sijbers, J., Jones, D., 2009. ExploreDTI: a graphical toolbox for processing, analyzing, and visualizing diffusion MR data, in: 17th annual meeting of intl soc mag reson med, International Society for Magnetic Resonance in Medicine Berkeley, CA, USA. p. 3537.
- Ling, J.M., Pena, A., Yeo, R.A., Merideth, F.L., Klimaj, S., Gasparovic, C., Mayer, A.R., 2012. Biomarkers of increased diffusion anisotropy in semi-acute mild traumatic brain injury: a longitudinal perspective. *Brain* 135, 1281–1292.

## BIBLIOGRAPHY

- Lister, J.P., Barnes, C.A., 2009. Neurobiological changes in the hippocampus during normative aging. *Archives of Neurology* 66, 829–833.
- Lohrke, J., Frisk, A.L., Frenzel, T., Schöckel, L., Rosenbruch, M., Jost, G., Lenhard, D.C., Sieber, M.A., Nischwitz, V., Küppers, A., et al., 2017. Histology and gadolinium distribution in the rodent brain after the administration of cumulative high doses of linear and macrocyclic gadolinium-based contrast agents. *Investigative radiology* 52, 324.
- Lombaert, H., Grady, L., Polimeni, J.R., Cheriet, F., 2012. FOCUSR: Feature oriented correspondence using spectral regularization-a method for accurate surface matching. *IEEE transactions on pattern analysis and machine intelligence* , 1.
- Ludwig, E., Klingler, J., 1956. The inner structure of the brain demonstrated on the basis of macroscopical preparations. *Atlas Cerebri Humani*. Brown Boston: Little .
- Maffei, C., Jovicich, J., De Benedictis, A., Corsini, F., Barbareschi, M., Chioffi, F., Sarubbo, S., 2018. Topography of the human acoustic radiation as revealed by ex vivo fibers microdissection and in vivo diffusion-based tractography. *Brain Structure and Function* 223, 449–459.
- Maier-Hein, K.H., Neher, P.F., Houde, J.C., Côté, M.A., Garyfallidis, E., Zhong, J., Chamberland, M., Yeh, F.C., Lin, Y.C., Ji, Q., et al., 2017. The challenge of mapping the human connectome based on diffusion tractography. *Nature communications* 8, 1349.
- Makris, N., Kennedy, D.N., McInerney, S., Sorensen, A.G., Wang, R., Caviness Jr, V.S., Pandya, D.N., 2004. Segmentation of subcomponents within the superior longitudinal fascicle in humans: a quantitative, in vivo, DT-MRI study. *Cerebral cortex* 15, 854–869.
- Mandonnet, E., Sarubbo, S., Petit, L., 2018. The nomenclature of human white matter association pathways: Proposal for a systematic taxonomic anatomical classification. *Frontiers in Neuroanatomy* 12, 94.
- Mangin, J.F., Fillard, P., Cointepas, Y., Le Bihan, D., Frouin, V., Poupon, C., 2013. Toward global tractography. *Neuroimage* 80, 290–296.
- Mansfield, P., 1977. Multi-planar image formation using NMR spin echoes. *Journal of Physics C: Solid State Physics* 10, L55.

## BIBLIOGRAPHY

- Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., Coffey, C., Kiebertz, K., Flagg, E., Chowdhury, S., et al., 2011. The parkinson progression marker initiative (PPMI). *Progress in neurobiology* 95, 629–635.
- Mars, R.B., Jbabdi, S., Sallet, J., O'Reilly, J.X., Croxson, P.L., Olivier, E., Noonan, M.P., Bergmann, C., Mitchell, A.S., Baxter, M.G., et al., 2011. Diffusion-weighted imaging tractography-based parcellation of the human parietal cortex and comparison with human and macaque resting-state functional connectivity. *Journal of Neuroscience* 31, 4087–4100.
- Masson, H.L., Kang, H.m., Petit, L., Wallraven, C., 2018. Neuroanatomical correlates of haptic object processing: combined evidence from tractography and functional neuroimaging. *Brain Structure and Function* 223, 619–633.
- Mazoyer, B., Mellet, E., Perchey, G., Zago, L., Crivello, F., Jobard, G., Delcroix, N., Vigneau, M., Leroux, G., Petit, L., et al., 2016. BIL&GIN: a neuroimaging, cognitive, behavioral, and genetic database for the study of human brain lateralization. *Neuroimage* 124, 1225–1231.
- Mazzocchi, F., Vignolo, L.A., 1979. Localisation of lesions in aphasia: clinical-CT scan correlations in stroke patients. *Cortex* 15, 627–653.
- Meola, A., Comert, A., Yeh, F.C., Stefaneanu, L., Fernandez-Miranda, J.C., 2015. The controversial existence of the human superior fronto-occipital fasciculus: Connectome-based tractographic study with microdissection validation. *Human brain mapping* 36, 4964–4971.
- Meyer, C.C., Meyer, O.N., 1973. Methods for brain autoradiography. *Physiology & behavior* 11, 265–267.
- Mole, J.P., Subramanian, L., Bracht, T., Morris, H., Metzler-Baddeley, C., Linden, D.E., 2016. Increased fractional anisotropy in the motor tracts of Parkinson's disease suggests compensatory neuroplasticity or selective neurodegeneration. *European radiology* 26, 3327–3335.
- Mollink, J., Kleinnijenhuis, M., van Walsum, A.M.v.C., Sotiropoulos, S.N., Cottaar, M., Mirfin, C., Heinrich, M.P., Jenkinson, M., Pallebage-Gamarallage, M., Ansorge, O., et al.,

## BIBLIOGRAPHY

2017. Evaluating fibre orientation dispersion in white matter: comparison of diffusion MRI, histology and polarized light imaging. *Neuroimage* 157, 561–574.
- Morawski, M., Kirilina, E., Scherf, N., Jäger, C., Reimann, K., Trampel, R., Gavriilidis, F., Geyer, S., Biedermann, B., Arendt, T., et al., 2018. Developing 3D microscopy with CLARITY on human brain tissue: Towards a tool for informing and validating MRI-based histology. *Neuroimage* 182, 417–428.
- Mori, S., Crain, B.J., Chacko, V.P., Van Zijl, P.C., 1999. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 45, 265–269.
- Mori, S., Wakana, S., Van Zijl, P.C., Nagae-Poetscher, L., 2005. MRI atlas of human white matter. Elsevier.
- Mori, S., van Zijl, P., 2002. Fiber tracking: principles and strategies—a technical review. *NMR in Biomedicine* 15, 468–480.
- Moseley, M., Cohen, Y., Mintorovitch, J., Chileuitt, L., Shimizu, H., Kucharczyk, J., Wendland, M., Weinstein, P., 1990. Early detection of regional cerebral ischemia in cats: comparison of diffusion-and T2-weighted MRI and spectroscopy. *Magnetic resonance in medicine* 14, 330–346.
- Moyer, D.C., Thompson, P., Ver Steeg, G., 2018. Measures of Tractography Convergence, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 295–307.
- Neher, P.F., Stieltjes, B., Reisert, M., Reicht, I., Meinzer, H.P., Fritzsche, K.H., 2012. MITK global tractography, in: *Medical Imaging 2012: Image Processing*, International Society for Optics and Photonics. p. 83144D.
- Nilsson, M., Englund, E., Szczepankiewicz, F., van Westen, D., Sundgren, P.C., 2018. Imaging brain tumour microstructure. *Neuroimage* 182, 232–250.
- O’donnell, L.J., Golby, A.J., Westin, C.F., 2013. Fiber clustering versus the parcellation-based connectome. *NeuroImage* 80, 283–289.

## BIBLIOGRAPHY

- O'Donnell, L.J., Pasternak, O., 2015. Does diffusion MRI tell us anything about the white matter? An overview of methods and pitfalls. *Schizophrenia research* 161, 133–141.
- O'Donnell, L.J., Suter, Y., Rigolo, L., Kahali, P., Zhang, F., Norton, I., Albi, A., Olubiyi, O., Meola, A., Essayed, W.I., et al., 2017. Automated white matter fiber tract identification in patients with brain tumors. *NeuroImage: Clinical* 13, 138–153.
- Panesar, S., Fernandez-Miranda, J., 2019. Commentary: The Nomenclature of Human White Matter Association Pathways: Proposal for a Systematic Taxonomic Anatomical Classification. *Frontiers in Neuroanatomy* 13, 61.
- Pasternak, O., Sochen, N., Gur, Y., Intrator, N., Assaf, Y., 2009. Free water elimination and mapping from diffusion MRI. *Magnetic resonance in medicine* 62, 717–730.
- Peled, S., Friman, O., Jolesz, F., Westin, C.F., 2006. Geometrically constrained two-tensor model for crossing tracts in DWI. *Magnetic resonance imaging* 24, 1263–1270.
- Pestilli, F., Yeatman, J.D., Rokem, A., Kay, K.N., Wandell, B.A., 2014. Evaluation and statistical inference for human connectomes. *Nature methods* 11, 1058.
- Peters, C.A., 2001. Statistics for analysis of experimental data. *Environmental engineering processes laboratory manual* , 1–25.
- Petersen, R.C., Aisen, P., Beckett, L., Donohue, M., Gamst, A., Harvey, D., Jack, C., Jagust, W., Shaw, L., Toga, A., et al., 2010. Alzheimer's disease Neuroimaging Initiative (ADNI) clinical characterization. *Neurology* 74, 201–209.
- Piccinini, F., Tesei, A., Paganelli, G., Zoli, W., Bevilacqua, A., 2014. Improving reliability of live/dead cell counting through automated image mosaicing. *Computer methods and programs in biomedicine* 117, 448–463.
- Pierpaoli, C., Basser, P.J., 1996. Toward a quantitative assessment of diffusion anisotropy. *Magnetic resonance in Medicine* 36, 893–906.
- Pipitone, J., Park, M.T.M., Winterburn, J., Lett, T.A., Lerch, J.P., Pruessner, J.C., Lepage, M., Voineskos, A.N., Chakravarty, M.M., Initiative, A.D.N., et al., 2014. Multi-atlas segmentation of the whole hippocampus and subfields using multiple automatically generated templates. *Neuroimage* 101, 494–512.

## BIBLIOGRAPHY

- Portegies, J.M., Fick, R.H.J., Sanguinetti, G.R., Meesters, S.P., Girard, G., Duits, R., 2015. Improving fiber alignment in HARDI by combining contextual PDE flow with constrained spherical deconvolution. *PloS one* 10, e0138122.
- Poulin, P., Cote, M.A., Houde, J.C., Petit, L., Neher, P.F., Maier-Hein, K.H., Larochelle, H., Descoteaux, M., 2017. Learn to track: Deep learning for tractography, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 540–547.
- Poulin, P., Jorgens, D., Jodoin, P.M., Descoteaux, M., 2019. Tractography and machine learning: Current state and open challenges. *Magnetic Resonance Imaging* .
- Poupon, C., Clark, C., Frouin, V., Regis, J., Bloch, I., Le Bihan, D., Mangin, J.F., 2000. Regularization of diffusion-based direction maps for the tracking of brain white matter fascicles. *Neuroimage* 12, 184–195.
- Presseau, C., Jodoin, P.M., Houde, J.C., Descoteaux, M., 2015. A new compression format for fiber tracking datasets. *NeuroImage* 109, 73–83.
- Pujol, S., Wells, W., Pierpaoli, C., Brun, C., Gee, J., Cheng, G., Vemuri, B., Commowick, O., Prima, S., Stamm, A., et al., 2015. The DTI challenge: toward standardized evaluation of diffusion tensor imaging tractography for neurosurgery. *Journal of Neuroimaging* 25, 875–882.
- Purcell, E.M., Torrey, H.C., Pound, R.V., 1946. Resonance absorption by nuclear magnetic moments in a solid. *Physical review* 69, 37.
- Qi, S., Meesters, S., Nicolay, K., ter Haar Romeny, B.M., Ossenblok, P., 2015. The influence of construction methodology on structural brain network measures: A review. *Journal of neuroscience methods* 253, 170–182.
- Raffelt, D.A., Tournier, J.D., Smith, R.E., Vaughan, D.N., Jackson, G., Ridgway, G.R., Connelly, A., 2017. Investigating white matter fibre density and morphology using fixel-based analysis. *Neuroimage* 144, 58–73.
- Reckfort, J., Wiese, H., Pietrzyk, U., Zilles, K., Amunts, K., Axer, M., 2015. A multiscale approach for the reconstruction of the fiber architecture of the human brain based on 3D-PLI. *Frontiers in neuroanatomy* 9, 118.

## BIBLIOGRAPHY

- Reisert, M., Kiselev, V.G., 2011. Fiber continuity: an anisotropic prior for ODF estimation. *IEEE transactions on medical imaging* 30, 1274–1283.
- Reisert, M., Mader, I., Anastasopoulos, C., Weigel, M., Schnell, S., Kiselev, V., 2011. Global fiber reconstruction becomes practical. *Neuroimage* 54, 955–962.
- Reitz, C., Brickman, A.M., Brown, T.R., Manly, J., DeCarli, C., Small, S.A., Mayeux, R., 2009. Linking hippocampal structure and function to memory performance in an aging population. *Archives of neurology* 66, 1385–1392.
- Renauld, E., Descoteaux, M., Bernier, M., Garyfallidis, E., Whittingstall, K., 2016. Semi-Automatic Segmentation of Optic Radiations and LGN, and Their Relationship to EEG Alpha Waves. *PloS one* 11, e0156436.
- Reveley, C., Seth, A.K., Pierpaoli, C., Silva, A.C., Yu, D., Saunders, R.C., Leopold, D.A., Frank, Q.Y., 2015. Superficial white matter fiber systems impede detection of long-range cortical connections in diffusion MR tractography. *Proceedings of the National Academy of Sciences* 112, E2820–E2828.
- Rheault, F., De Benedictis, A., Daducci, A., Maffei, C., Tax, C.M., Romascano, D., Caverzasi, E., Morency, F.C., Corrivetti, F., Pestilli, F., et al., 2020. Tractostorm: The what, why, and how of tractography dissection reproducibility. *Human Brain Mapping* .
- Rheault, F., Houde, J.C., Descoteaux, M., 2017. Visualization, Interaction and Tractometry: Dealing with Millions of Streamlines from Diffusion MRI Tractography. *Frontiers in neuroinformatics* 11, 42.
- Rheault, F., Houde, J.C., Goyette, N., Morency, F., Descoteaux, M., 2016. MI-Brain, a software to handle tractograms and perform interactive virtual dissection, in: *ISMRM Diffusion study group workshop*, Lisbon.
- Rheault, F., Roy, M., Cunnane, S., Descoteaux, M., 2018a. Bundle-specific fornix reconstruction for dual-tracer PET-tractometry. *bioRxiv* , 423459.
- Rheault, F., St-Onge, E., Sidhu, J., Chenot, Q., Petit, L., Descoteaux, M., 2018b. Bundle-Specific Tractography, in: *Computational Diffusion MRI*. Springer, pp. 129–139.

## BIBLIOGRAPHY

- Rheault, F., St-Onge, E., Sidhu, J., Maier-Hein, K., Tzourio-Mazoyer, N., Petit, L., Descoteaux, M., 2019. Bundle-specific tractography with incorporated anatomical and orientational priors. *NeuroImage* 186, 382–398.
- ROBINSON, R.G., KUBOS, K.L., STARR, L.B., RAO, K., PRICE, T.R., 1984. Mood disorders in stroke patients: importance of location of lesion. *Brain* 107, 81–93.
- Roldan-Valadez, E., Lopez-Mejia, M., 2014. Current concepts on magnetic resonance imaging (MRI) perfusion-diffusion assessment in acute ischaemic stroke: a review & an update for the clinicians. *The Indian journal of medical research* 140, 717.
- Rozanski, V.E., da Silva, N.M., Ahmadi, S.A., Mehrkens, J., da Silva Cunha, J., Houde, J.C., Vollmar, C., Bötzel, K., Descoteaux, M., 2017. The role of the pallidothalamic fibre tracts in deep brain stimulation for dystonia: a diffusion MRI tractography study. *Human brain mapping* 38, 1224–1232.
- Rushworth, M., Behrens, T., Johansen-Berg, H., 2005. Connection patterns distinguish 3 regions of human parietal cortex. *Cerebral cortex* 16, 1418–1430.
- Sakaie, K.E., Lowe, M.J., 2007. An objective method for regularization of fiber orientation distributions derived from diffusion-weighted MRI. *NeuroImage* 34, 169–176.
- Saleeba, C., Dempsey, B.R., Le, S., Goodchild, A.K., McMullan, S., 2019. A student’s guide to neural circuit tracing. *Frontiers in Neuroscience* 13, 897.
- Sarubbo, S., Petit, L., De Benedictis, A., Chioffi, F., Ptito, M., Dyrby, T.B., 2019. Uncovering the inferior fronto-occipital fascicle and its topological organization in non-human primates: the missing connection for language evolution. *Brain Structure and Function* 224, 1553–1567.
- Schiavi, S., Barakovic, M., Ocampo-Pineda, M., Descoteaux, M., Thiran, J.P., Daducci, A., 2019. Reducing false positives in tractography with microstructural and anatomical priors. *bioRxiv* , 608349.
- Schilling, K., Gao, Y., Janve, V., Stepniewska, I., Landman, B.A., Anderson, A.W., 2018a. Confirmation of a gyral bias in diffusion MRI fiber tractography. *Human brain mapping* 39, 1449–1466.



## BIBLIOGRAPHY

- Schilling, K.G., Nath, V., Hansen, C., Parvathaneni, P., Blaber, J., Gao, Y., Neher, P., Aydogan, D.B., Shi, Y., Ocampo-Pineda, M., et al., 2018b. Limits to anatomical accuracy of diffusion tractography using modern approaches. *bioRxiv* , 392571.
- Schmahmann, J., Pandya, D., 2006. Fiber pathways of the brain. OUP USA.
- Schmahmann, J.D., Pandya, D.N., Wang, R., Dai, G., D’Arceuil, H.E., de Crespigny, A.J., Wedeen, V.J., 2007. Association fibre pathways of the brain: parallel observations from diffusion spectrum imaging and autoradiography. *Brain* 130, 630–653.
- Smith, R.E., Tournier, J.D., Calamante, F., Connelly, A., 2012. Anatomically-constrained tractography: improved diffusion MRI streamlines tractography through effective use of anatomical information. *Neuroimage* 62, 1924–1938.
- Smith, R.E., Tournier, J.D., Calamante, F., Connelly, A., 2015a. SIFT2: Enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography. *Neuroimage* 119, 338–351.
- Smith, R.E., Tournier, J.D., Calamante, F., Connelly, A., 2015b. The effects of SIFT on the reproducibility and biological accuracy of the structural connectome. *Neuroimage* 104, 253–265.
- Song, J.W., Mitchell, P.D., Kolasinski, J., Ellen Grant, P., Galaburda, A.M., Takahashi, E., 2014. Asymmetry of white matter pathways in developing human brains. *Cerebral cortex* 25, 2883–2893.
- Sotiropoulos, S.N., Zalesky, A., 2019. Building connectomes using diffusion MRI: why, how and but. *NMR in Biomedicine* 32, e3752.
- Spence, R.D., Kurth, F., Itoh, N., Mongerson, C.R., Wailes, S.H., Peng, M.S., MacKenzie-Graham, A.J., 2014. Bringing CLARITY to gray matter atrophy. *Neuroimage* 101, 625–632.
- Sporns, O., Tononi, G., Kötter, R., 2005. The human connectome: a structural description of the human brain. *PLoS computational biology* 1, e42.
- Sprengers, M., Vonck, K., Carrette, E., Marson, A.G., Boon, P., 2017. Deep brain and cortical stimulation for epilepsy. *Cochrane Database of Systematic Reviews* .

## BIBLIOGRAPHY

- St-Onge, E., Daducci, A., Girard, G., Descoteaux, M., 2018. Surface-enhanced tractography (SET). *NeuroImage* 169, 524–539.
- Storsve, A.B., Fjell, A.M., Yendiki, A., Walhovd, K.B., 2016. Longitudinal changes in white matter tract integrity across the adult lifespan and its relation to cortical thinning. *PLoS One* 11, e0156770.
- Subbanna, N.K., Rajasheka, D., Cheng, B., Thomalla, G., Fiehler, J., Arbel, T., Forkert, N.D., 2019. Stroke lesion segmentation in FLAIR MRI datasets using customized Markov random fields. *Frontiers in neurology* 10, 541.
- Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al., 2015. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine* 12, e1001779.
- Suess, O., Suess, S., Brock, M., Kombos, T., 2006. Intraoperative electrocortical stimulation of Brodman area 4: a 10-year analysis of 255 cases. *Head & face medicine* 2, 20.
- Swanson, L., 2000. A history of neuroanatomical mapping, in: *Brain Mapping: The Systems*. Elsevier, pp. 77–109.
- Szelényi, A., Bello, L., Duffau, H., Fava, E., Feigl, G.C., Galanda, M., Neuloh, G., Signorelli, F., Sala, F., 2010. Intraoperative electrical stimulation in awake craniotomy: methodological aspects of current practice. *Neurosurgical focus* 28, E7.
- Tax, C.M., Jeurissen, B., Vos, S.B., Viergever, M.A., Leemans, A., 2014. Recursive calibration of the fiber response function for spherical deconvolution of diffusion MRI data. *Neuroimage* 86, 67–80.
- Teillac, A., Beaujoin, J., Poupon, F., Mangin, J.F., Poupon, C., 2017. A novel anatomically-constrained global tractography approach to monitor sharp turns in Gyri, in: *International Conference on Medical Image Computing and Computer-assisted Intervention*, Springer. pp. 532–539.
- Theaud, G., Houde, J.C., Boré, A., Rheault, F., Morency, F., Descoteaux, M., 2019. TractoFlow: A robust, efficient and reproducible diffusion MRI pipeline leveraging Nextflow & Singularity. *bioRxiv* , 631952.

## BIBLIOGRAPHY

- Thirion, B., Varoquaux, G., Dohmatob, E., Poline, J.B., 2014. Which fMRI clustering gives good brain parcellations? *Frontiers in neuroscience* 8, 167.
- Thomas, C., Frank, Q.Y., Irfanoglu, M.O., Modi, P., Saleem, K.S., Leopold, D.A., Pierpaoli, C., 2014. Anatomical accuracy of brain connections derived from diffusion MRI tractography is inherently limited. *Proceedings of the National Academy of Sciences* 111, 16574–16579.
- Thomas Yeo, B., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., et al., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology* 106, 1125–1165.
- Tittgemeyer, M., Rigoux, L., Knösche, T.R., 2018. Cortical parcellation based on structural connectivity: a case for generative models. *Neuroimage* 173, 592–603.
- Tournier, J., Calamante, F., Connelly, A., et al., 2012. MRtrix: diffusion tractography in crossing fiber regions. *International Journal of Imaging Systems and Technology* 22, 53–66.
- Tournier, J.D., Calamante, F., Connelly, A., 2007. Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. *Neuroimage* 35, 1459–1472.
- Tournier, J.D., Yeh, C.H., Calamante, F., Cho, K.H., Connelly, A., Lin, C.P., 2008. Resolving crossing fibres using constrained spherical deconvolution: validation using diffusion-weighted imaging phantom data. *Neuroimage* 42, 617–625.
- Tuch, D.S., 2004. Q-ball imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 52, 1358–1372.
- Tuch, D.S., Reese, T.G., Wiegell, M.R., Makris, N., Belliveau, J.W., Wedeen, V.J., 2002. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 48, 577–582.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using

## BIBLIOGRAPHY

- a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289.
- Valk, S.L., Di Martino, A., Milham, M.P., Bernhardt, B.C., 2015. Multicenter mapping of structural network alterations in autism. *Human brain mapping* 36, 2364–2373.
- Van Essen, D.C., Jbabdi, S., Sotiropoulos, S.N., Chen, C., Dikranian, K., Coalson, T., Harwell, J., Behrens, T.E., Glasser, M.F., 2014. Mapping connections in humans and non-human primates: aspirations and challenges for diffusion imaging, in: *Diffusion MRI*. Elsevier, pp. 337–358.
- Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, W.M.H., et al., 2013. The WU-Minn human connectome project: an overview. *Neuroimage* 80, 62–79.
- Van Everdingen, K., Van der Grond, J., Kappelle, L., Ramos, L., Mali, W., 1998. Diffusion-weighted magnetic resonance imaging in acute stroke. *Stroke* 29, 1783–1790.
- Verleysen, M., François, D., 2005. The curse of dimensionality in data mining and time series prediction, in: *International Work-Conference on Artificial Neural Networks*, Springer. pp. 758–770.
- Voineskos, A.N., O’donnell, L.J., Lobaugh, N.J., Markant, D., Ameis, S.H., Niethammer, M., Mulsant, B.H., Pollock, B.G., Kennedy, J.L., Westin, C.F., et al., 2009. Quantitative examination of a novel clustering method using magnetic resonance diffusion tensor tractography. *Neuroimage* 45, 370–376.
- Wakana, S., Caprihan, A., Panzenboeck, M.M., Fallon, J.H., Perry, M., Gollub, R.L., Hua, K., Zhang, J., Jiang, H., Dubey, P., et al., 2007. Reproducibility of quantitative tractography methods applied to cerebral white matter. *Neuroimage* 36, 630–644.
- Wang, H., Lenglet, C., Akkin, T., 2015. Structure tensor analysis of serial optical coherence scanner images for mapping fiber orientations and tractography in the brain. *Journal of biomedical optics* 20, 036003.
- Wang, R., Benner, T., Sorensen, A.G., Wedeen, V.J., 2007. Diffusion toolkit: a software package for diffusion imaging data processing and tractography, in: *Proc Intl Soc Mag Reson Med*, Berlin.

## BIBLIOGRAPHY

- Wang, X., Grimson, W.E.L., Westin, C.F., 2011. Tractography segmentation using a hierarchical Dirichlet processes mixture model. *NeuroImage* 54, 290–302.
- Warach, S., Chien, D., Li, W., Ronthal, M., Edelman, R., 1992. Fast magnetic resonance diffusion-weighted imaging of acute human stroke. *Neurology* 42, 1717–1717.
- Wassermann, D., Makris, N., Rathi, Y., Shenton, M., Kikinis, R., Kubicki, M., Westin, C.F., 2016. The white matter query language: a novel approach for describing human white matter anatomy. *Brain Structure and Function* 221, 4705–4721.
- Wasserthal, J., Neher, P., Maier-Hein, K.H., 2018a. Tractseg-fast and accurate white matter tract segmentation. *NeuroImage* 183, 239–253.
- Wasserthal, J., Neher, P.F., Maier-Hein, K.H., 2018b. Tract orientation mapping for bundle-specific tractography. *arXiv preprint arXiv:1806.05580* .
- Wedeen, V.J., Wang, R., Schmahmann, J.D., Benner, T., Tseng, W., Dai, G., Pandya, D., Hagmann, P., D’Arceuil, H., de Crespigny, A.J., 2008. Diffusion spectrum magnetic resonance imaging (DSI) tractography of crossing fibers. *Neuroimage* 41, 1267–1277.
- Wei, K., Cieslak, M., Greene, C., Grafton, S.T., Carlson, J.M., 2017. Sensitivity analysis of human brain structural network construction. *Network Neuroscience* 1, 446–467.
- Wender, R., Brown, A.M., Fern, R., Swanson, R.A., Farrell, K., Ransom, B.R., 2000. Astrocytic glycogen influences axon function and survival during glucose deprivation in central white matter. *Journal of Neuroscience* 20, 6804–6810.
- Williams, T.H., Gluhbegovic, N., Jew, J.Y., 1999. The human brain: dissections of the real brain. University of Iowa.
- Wisse, L.E., Daugherty, A.M., Olsen, R.K., Berron, D., Carr, V.A., Stark, C.E., Amaral, R.S., Amunts, K., Augustinack, J.C., Bender, A.R., et al., 2017. A harmonized segmentation protocol for hippocampal and parahippocampal subregions: Why do we need one and what are the key goals? *Hippocampus* 27, 3–11.
- Witelson, S.F., 1985. The brain connection: the corpus callosum is larger in left-handers. *Science* 229, 665–668.

## BIBLIOGRAPHY

- Woolrich, M.W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., Smith, S.M., 2009. Bayesian analysis of neuroimaging data in FSL. *Neuroimage* 45, S173–S186.
- Yeatman, J.D., Dougherty, R.F., Myall, N.J., Wandell, B.A., Feldman, H.M., 2012. Tract profiles of white matter properties: automating fiber-tract quantification. *PloS one* 7, e49790.
- Yeatman, J.D., Richie-Halford, A., Smith, J.K., Keshavan, A., Rokem, A., 2018. A browser-based tool for visualization and analysis of diffusion MRI data. *Nature communications* 9, 940.
- Yeh, C.H., Smith, R.E., Dhollander, T., Calamante, F., Connelly, A., 2019. Connectomes from streamlines tractography: Assigning streamlines to brain parcellations is not trivial but highly consequential. *NeuroImage* 199, 160–171.
- Yeh, C.H., Smith, R.E., Liang, X., Calamante, F., 2018a. Investigating the streamline count required for reproducible structural connectome construction across a range of brain parcellation resolutions, in: *Proc ISMRM*, p. 1558.
- Yeh, F.C., Panesar, S., Fernandes, D., Meola, A., Yoshino, M., Fernandez-Miranda, J.C., Vettel, J.M., Verstynen, T., 2018b. Population-averaged atlas of the macroscale human structural connectome and its network topology. *NeuroImage* 178, 57–68.
- Yeh, F.C., Tseng, W.Y.I., 2011. NTU-90: a high angular resolution brain atlas constructed by q-space diffeomorphic reconstruction. *Neuroimage* 58, 91–99.
- Yendiki, A., Panneck, P., Srinivasan, P., Stevens, A., Zöllei, L., Augustinack, J., Wang, R., Salat, D., Ehrlich, S., Behrens, T., et al., 2011. Automated probabilistic reconstruction of white-matter pathways in health and disease using an atlas of the underlying anatomy. *Frontiers in neuroinformatics* 5, 23.
- Yeterian, E.H., Pandya, D.N., Tomaiuolo, F., Petrides, M., 2012. The cortical connectivity of the prefrontal cortex in the monkey brain. *Cortex* 48, 58–81.
- Yoo, S.W., Guevara, P., Jeong, Y., Yoo, K., Shin, J.S., Mangin, J.F., Seong, J.K., 2015. An example-based multi-atlas approach to automatic labeling of white matter tracts. *PloS one* 10, e0133337.

## BIBLIOGRAPHY

- Záborszky, L., Wouterlood, F.G., Lanciego, J.L., 2006. Neuroanatomical tract-tracing: molecules, neurons, and systems. Springer Science & Business Media.
- Zhang, F., Wu, W., Ning, L., McAnulty, G., Waber, D., Gagoski, B., Sarill, K., Hamoda, H.M., Song, Y., Cai, W., et al., 2018a. Suprathreshold fiber cluster statistics: Leveraging white matter geometry to enhance tractography statistical analysis. *NeuroImage* .
- Zhang, Y., Zhang, J., Oishi, K., Faria, A.V., Jiang, H., Li, X., Akhter, K., Rosa-Neto, P., Pike, G.B., Evans, A., et al., 2010. Atlas-guided tract reconstruction for automated and comprehensive examination of the white matter anatomy. *Neuroimage* 52, 1289–1301.
- Zhang, Z., Descoteaux, M., Zhang, J., Girard, G., Chamberland, M., Dunson, D., Srivastava, A., Zhu, H., 2018b. Mapping population-based structural connectomes. *NeuroImage* 172, 130–145.

“The world is not in your books and maps, it is out there.”

---

— Gandalf, *The Hobbit*.